Value Stream Financial Modeling for Improved Production Decision Making

by

Christopher Warren Hopkins

B.S. Electrical and Computer Engineering, Cornell University, 2005

Submitted to the MIT Sloan School of Management and the Engineering Systems Division in Partial Fulfillment of the Requirements for the Degrees of

Master of Business Administration
AND
Master of Science in Engineering Systems

In conjunction with the Leaders for Global Operations Program at the

Massachusetts Institute of Technology
June 2011

Signature of Author _____
May 6, 2011
Engineering Systems Division, MIT Sloan School of Management

Certified by _____
Roy Welsch, Thesis Supervisor
Professor of Statistics and Management Science and Engineering Systems

Certified by _____
David Simchi-Levi, Thesis Supervisor
Professor of Engineering Systems and Civil & Environmental Engineering

Accepted by _____
Nancy Leveson, Chair, Engineering Systems Division Education Committee
Professor Aeronautics and Astronautics and Engineering Systems Division

Accepted by _____
Debbie Berechman, Executive Director of MBA Program
MIT Sloan School of Management

This page intentionally left blank.

Value Stream Financial Modeling for Improved Production Decision Making

by

Christopher Warren Hopkins

Submitted to the MIT Sloan School of Management and the Engineering Systems Division
on May 6, 2011 in Partial Fulfillment of the Requirements for the Degrees of

Master of Business Administration
AND
Master of Science in Engineering Systems

ABSTRACT

Understanding the overall impact of a decision in a manufacturing system can be challenging given the complex production and financial structures in today's companies. While knowing the direct result of a local change may be easy, anticipating the real impact to the rest of the business can be difficult. Nonetheless, managers are faced with this dilemma on a regular basis as they try to support the larger organization, taking appropriate actions as best they can.

Based on a project at Novartis Vaccines and Diagnostics for an influenza vaccine, this thesis helps address some of the key questions managers face. It discusses a technique for more accurately determining the implications of these common manufacturing decisions:

- How much should be spent to improve a particular component?
- What are the impacts of expanding into new markets?
- Which parameters in the factory most deserve managerial attention?
- What are the appropriate tradeoffs to make when deciding on materials purchasing?

Using concepts from throughput accounting, a model is developed from a detailed cost structure analysis, linking the financial and production aspects of the system. Whenever a parameter is changed, the model simulates how the rest of the system would perform through a linear program that replicates the production scheduling process. Thus, a manager is able to experiment with the tool in order to observe the overall impact of the change being considered and levy a decision based on the anticipated costs and benefits projected by the model. As a result, managers can distribute resources in a more efficient manner and align decision making throughout the organization.

This thesis discusses the modeling approach, historical validation and initial insights for the current system. It also covers techniques for future applications and identifies the underlying organizational challenges that must be addressed to achieve a global optimum.

Thesis Supervisor: Roy Welsch
Title: Professor of Statistics and Management Science and Engineering Systems

Thesis Supervisor: David Simchi-Levi
Title: Professor of Engineering Systems and Civil & Environmental Engineering

This page intentionally left blank.

ACKNOWLEDGMENTS

I am tremendously thankful, not only for those people that have helped make this project a success, but also for everybody that has supported me throughout my journey.

Thank you to the people of Novartis who put their trust in me to execute this project and the help to make it successful. I would specifically like to thank my supervisor, Ian Dick, mentor, Brian Orman and sponsor, John Sullivan for all of their help. Additionally, Mark Hyde and David Murphy went above and beyond to share their knowledge with me. I had a terrific learning experience and I now have a great appreciation for the work that you all perform.

My MIT professors, Roy Welsch and David Simchi-Levi, were terrific resources who provided valuable insights along the way. Thank you for being my advisors.

To Don Rosenfield, faculty, staff and my 46 new friends in the Leaders for Global Operations program, thank you for welcoming me into this amazing community. The past two years has been a life-changing experience and I cannot imagine having spent my time doing anything else.

Finally, thank you to my parents and brother for their love and support. I would not have been able to get this far without you and I love you all very much.

This page intentionally left blank.

TABLE OF CONTENTS

TABLE OF FIGURES

CHAPTER 1: INTRODUCTION AND THESIS OVERVIEW

1.1 Problem Description and Motivation

As organizations evolve to be more and more complex, it has become increasingly challenging for managers to understand the overall impact of their decisions. Yet many have identified the ability to assess how local decisions affect the company as one of the foundations for running a business effectively (Goldratt 1990, 55; Corbett 1998, 142; Deming 2000, 87; Womack and Jones 2003, 26).

Unfortunately, the tools that should exist to help managers in this way have been unable to keep up with the evolution of many industries (Garg et al. 2003, 3). The leadership team at Novartis Vaccines & Diagnostics which manages the flu vaccine, Fluvirin, recognizes this issue. It has become too difficult to trust that they understand the real consequences of various managerial actions by using existing techniques. The sophisticated financial accounting systems that exist for legal reporting purposes simply do not provide the relevant information to support effective decision making.

In addition to the common challenges that managers face about understanding the impact of their decisions, several other factors complicate the environment in which the Fluvirin team operates. First, the influenza vaccine has seasonal demand as well as seasonal production. Patients only receive the vaccine during a few months of the year making sales and production leveling challenging. Additionally, the flu vaccine must change each year in response to the evolving flu strains present in nature. Production therefore operates in a "fast clock-speed" setting where managers have very little time to make each decision (Fine 1998, 19). Secondly, the significant amount of indirect work common to the biopharmaceutical industry (often associated with the quality testing for patient safety) is not well characterized and therefore concerns managers with respect to the real cost structure within production. Finally, and most importantly, flu vaccine is made in a multi-stage batch production process. The economics of this type of process are fundamentally different than for most other manufactured goods and require techniques for properly understanding the production system. Even some of the most basic concepts such as marginal costs become much more complicated in this mode of production.

This thesis is based on the work that was performed by the author over a six month period at the Liverpool, U.K. site of Novartis for the Fluvirin product. Many of the concepts shared are intended for a general audience, but are discussed in the context of the Fluvirin project. It attempts to assist in decision making for several of the key types of problems that the management team faces on a regular basis, and in a way that accurately reflects the true economic impact of those cost/benefit decisions by combining concepts from business and systems engineering.

1.1.1 Capital Investments and Project Spending

First, managers identify the need to understand and quantify the benefits associated with making a change in order to determine if it is worth the cost of investment. The local impact of a change is often easily understood, e.g. the installation of a new piece of equipment might raise the number of units that can be processed by 10%, but understanding how much of an overall financial impact that would have and whether it would be worth the price is much harder to determine. A solution is needed to bring together all the relevant information in an easy to interpret way in order to help assist such decisions.

1.1.2 Production and Sales Planning

Vaccine sales require close collaboration with the production facilities to ensure demand can be met at the appropriate times and to determine if new contracts can and should be signed. Part of that process requires an understanding of the profitability of changing production quantity plans and entering new markets. Fluvirin strategic planners need a better information system for making these decisions.

1.1.3 Process Improvement Project Evaluation

Modern initiatives such as Total Quality Management, Six Sigma and Lean are based on the concept of continuous improvement. However, an unfocused approach of selecting the project that seems to be the most useful may not in fact have the most significant overall impact. Additionally, they often do not connect back into the organization through financial measurements, thereby making it hard to compare projects and gain support from other key groups in the organization (Goldratt 1990, 55). Lean gurus Rother and Shook provide great tools for improving systems and identify project prioritization as important, yet little is presented to help formally quantify where to focus resources (Rother and Shook 2003, 90). The Theory of Constraints (TOC) provides more insights into where efforts should be focused by examining the whole production system, but adaptation will be needed to handle the multi-stage batch production process. Through an improved understanding of the system, managers will have a method for quickly knowing what types of continuous improvement projects they should pursue and which types should be postponed.

1.1.4 Materials Purchasing

In vaccine production, the quality of the incoming materials can often have a significant impact on the process. Most notably, the chicken eggs used in the flu vaccine production, can directly affect how many doses can be produced in each batch. Developing an understanding of this connection to the overall financials will help guide the site buyers in the appropriate tradeoff between cost and quality.

1.2 Current and Future State Analysis

In order to understand the context of the various challenges management faced, it is necessary to analyze the current state of the Fluvirin system and then identify the desired future state. Understanding the current state identifies why the previously mentioned managerial questions are not already answered and why the system is not already operating optimally. The future state shows how the new system needs to not only help answer those questions, but also how to create a system in which the optimal situation is achievable.

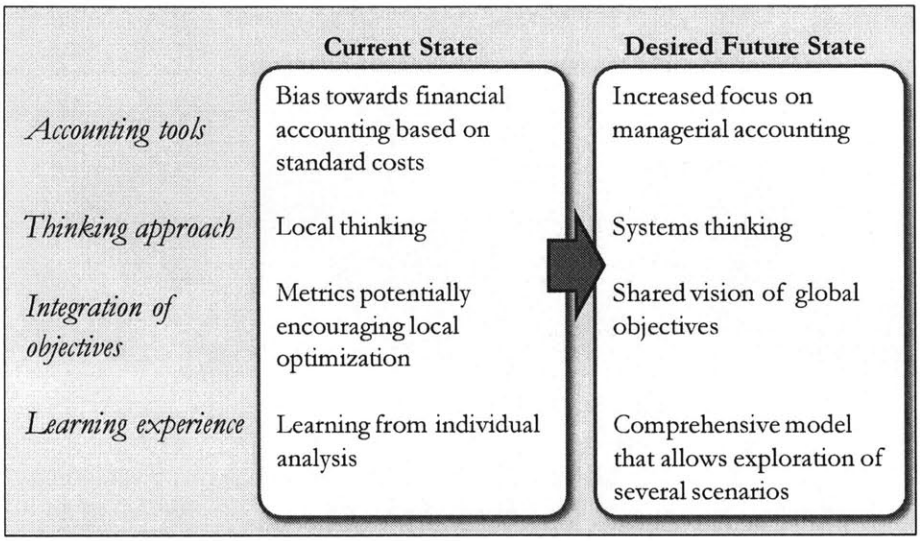

Figure 1: Current and Desired Future State

1.2.1 Accounting Tools

A brief discussion about the history of accounting is needed to understand why the existing accounting system at Novartis (and many other companies) does not provide managers with the necessary information to perform their jobs effectively. Most importantly, it is critical to first distinguish the difference between two different types of accounting:

- *Financial Accounting* is intended for external audiences (regulators, tax agencies, stockowners) for reporting purposes and is subject to governmental regulation (McNair and Vangermeersch 1998, 9)
- *Managerial Accounting* is intended for managers to inform decisions by linking their local actions to overall profitability (Corbett 1998, 142)

While there is some overlap in the two definitions and therefore some of the data used by both, the fundamental difference is that financial accounting is backwards looking to report on what has already happened, whereas managerial accounting is for future decisions. The important distinction is that managerial accounting must identify only the relevant information for the question at hand.

However, given that financial accounting is a "must have" because of the regulations for reporting and managerial accounting is a "nice to have", the perceived importance of financial accounting in many organizations has grown to overshadow its managerial counterpart (Huntzinger 2007, 19). In 1987, the growing concerns about how financial accounting techniques were negatively influencing managerial accounting were identified in *Relevance Lost*. "Today's management accounting information, driven by the procedures and cycle of the organization's financial reporting system, is too late, too aggregated, and too distorted to be relevant for managers' planning and control decisions" (Johnson and Kaplan 1987, 1).

The underlying reason why financial accounting often contains information that is not relevant for managers is that it is based on full (absorption) costing which aggregates all manufacturing costs and distributes them across products (McNair and Vangermeersch 1998, 9). Looking backwards from a historical perspective, this seems like a reasonable thing to do because it shows the average total cost, but when looking forward the numbers can be misleading. For example, based on a fully absorbed cost (also called standard cost) of $10 per part (which includes raw materials, labor, equipment, building lease, etc.) how much would making additional units really cost? Many of the values in the standard cost are fixed and would therefore not change in response to the higher production level. In reality, the real marginal cost of making an additional unit is less than the standard cost, but this fact is often obfuscated because standard costs are what is commonly reported. Using the standard cost would mislead a manager into thinking this increase would be more expensive than it really was and thus less profitable. Conversely, reducing the production volume would appear to save a substantial amount, but again since many of the costs are fixed, the real economic impact is that only the variable costs are avoided and the true savings are less significant. This distortion can lead managers into making suboptimal decisions.

Full costing was developed during a time in which materials and direct labor were the heavy majority of costs for a company and both were considered variable costs (Carnes and Hedin 2005, 29). At that time it would have been sensible to analyze a decision using full costing because that would have properly used the relevant costs. However, the economy has evolved to the point where labor (and many other costs) are fixed and are no longer relevant for consideration in short-term questions.

Finally, the cost accounting structure within many companies has developed in a way that it is no longer easy for managers to easily interpret the information because of the use of allocations (Stratton et al. 2009, 31). At Novartis, costs are first aggregated by the teams that provide support work (e.g. quality testing, security, and even the cafeteria) and allocated to the teams that provide the direct work. The allocations are then combined with the costs from the direct labor teams and then allocated again to the product at the appropriate process stage. This technique makes it tremendously difficult for managers to understand the underlying data and the real relationship of costs in the organization.

Several management experts have criticized cost accounting for the way it can mislead managers. Taiichi Ohno, originator of Toyota's lean concepts says, "... cost accounting sometimes leads top executives to faulty judgments" (Ohno 1988, 138). Even further, Eliyahu Goldratt even gave a conference presentation titled "Cost accounting is enemy number one of productivity" (Goldratt 1983).

Despite the clear shortcomings of full costing, or similar alternatives that use allocations, recent surveys have shown that a majority of manufacturing companies use accounting techniques that allocate costs (Stratton et al. 2009, 33). Novartis managers use standard costs for financial reporting purposes, but recognize that a substitute is needed in order to properly handle managerial decisions.

As an alternative to the techniques that use allocations, throughput accounting, based on the Theory of Constraints, provides a more realistic account of the short-term impact of managers' decisions. Rather than trying to break every part of the manufacturing system down through allocations like in absorption costing approaches, throughput accounting analyzes the system as a whole in order to understand the complete financial picture (Corbett 1998, 18). When a manager is contemplating a decision to make a change in one area, the two approaches become very distinct. Cost accounting treats every area in production as the same level of importance. In that world, saving an hour of processing time in one area is just as valuable as saving an hour somewhere else. Additionally, the allocation process creates internal money transfers that distort the real economic picture. Throughput accounting instead identifies how the hour savings impacts the overall production process (e.g. perhaps the savings in time now allows for more units to be produced, or maybe the modification has no impact on number of units that can be made). How the system would respond and thus what the overall financial impact would be depends on where the process change is that saves the hour. Also, only the real financial transactions are tracked rather than confusing the situation with artificial transfers of internal funds based on allocations (Corbett 1998, 116).

Throughput accounting has three main values: throughput, inventory (i.e. investment) and operating expenses (Goldratt 1999, 113). Throughput, defined as revenues minus truly variable costs,

will be the primary value that is focused on throughout this thesis because of the challenges associated with calculating it. To avoid confusion with the production term throughput [rate], meaning number of units processed in a period of time, [financial] throughput will often be referred to as the contribution margin, which is more consistent with common managerial accounting terminology. Investment and operating expenses are typically much more explicit when making a decision, so less emphasis will be placed on understanding those.

Multi-stage batch production requires a fundamentally different way of thinking about traditional concepts of costs and capacity which is why full costing even further deviates from what needs to be used in this setting. Though throughput accounting will serve as the basis for the approach used in the thesis for decision analysis, additional efforts to map the different types of costs for batch production will be needed in order to accurately model this system beyond what is normally used in throughput accounting.

For example, when additional units of a typical manufactured good are made, traditional thinking is that it will require an additional cost (i.e. the marginal cost), with additional capacity and for additional time. The same assumptions for multi-stage batch production, as in the case of flu vaccine production, do not always hold. The implications of what happens are dependent on how the additional units (i.e. doses) are produced. If the extra units are produced by making more batches, the total cost for those additional units is the cost of the additional batches and extra time will be required. However, if an improvement in the way the vaccine is grown in order to make additional doses from existing batches, no additional batch costs, capacity or batch processing time is needed. Traditional concepts are unable to make this important distinction because of their underlying assumptions. Therefore, instead of following a reductionist approach of breaking costs down to a cost per dose, we must instead pursue a systems approach which will consider the overall impact to the different types of costs in the system.

1.2.2 Thinking Approach

Like many large organizations, the team responsible for Fluvirin is separated across functional boundaries: finance, production, quality testing, supply chain, sales and marketing and so on. This structure creates a simple way to divide responsibility and develop expertise. The unfortunate drawback however, is that less attention is paid to the whole organization, even though it is the effort of the entire organization, not individual group, that creates the products. This leads to local thinking and the incorrect assumption that maximizing the individual components maximizes the system. Previous analytical approaches at Novartis struggled because they focused only on the particular pieces leading to local optimizations rather than a global optimization, which would improve the whole system by maximizing delivery to patients and profitability.

In order to achieve a globally optimum level, the system must be evaluated as a whole. Peter Senge, renowned for evangelizing systems thinking explains, "Living systems have integrity. Their character depends on the whole. The same is true for organizations; to understand the most challenging issues requires seeing the whole system that generates the issues" (Senge 2006, 66). Accordingly, this thesis will aim to understand the main groups of the organization in order to eventually understand their relationships within the system and to ultimately determine the most effective ways to improve the system.

Another aspect of systems is that the nature of their structure often generates a particular type of behavior (Meadows 1982, 98). For that reason, it is not sufficient to assume that behaviors will be changed for the long term without change to the underlying structure. Although the senior management of Novartis wants to maximize the overall system, this creates an underlying tension with the traditional desire to assess performance at group levels. Deming highlights the issue, "The obligation of any component is to contribute the best to the system, not to maximize its own production, profit or sales, nor any other competitive measure. Some components may operate at a loss to themselves in order to optimize the whole system..." (Deming 1993, 97). Without addressing this issue, it is likely that managers will have to make decisions that require tradeoffs between looking good on local metrics and supporting the organization's global objective.

1.2.3 Integration of Objectives

Deviations from a globally optimal solution can occur because of the different ways local evaluations are made. First, the financial responsibility for many managers in a production setting often only focuses on costs and ignores the more important financial measurement of profitability. Secondly, as was mentioned, separating functional groups prevents an understanding of the impact to the system. Developing a financial model using throughput accounting will help address both issues.

Cost centers are considered the main financial entity that managers are responsible for at Novartis. However, this emphasis improperly biases decisions toward reducing costs (in the assumption that minimum costs translates to maximum profits) instead of finding the optimal balance between costs and revenues. Additionally, the simplicity in calculating cost reductions makes them preferable to projects aimed at increasing revenue, regardless of which has the greater benefit. For example, suppose a manager has limited resources and is only able to complete one project. The first project reduces the amount of defects in the group by 1%, which could potentially increase production and sales. Understanding the total economic benefit is hard to do for someone in a functional role with a limited view of the rest of the organization. Meanwhile, the second project has identified that changing to a new

type of cleaning chemical will save 1% of the cleaning costs. In this case, computing the financial savings is straightforward. History of where efforts have been placed seems to suggest that the second project would be pursued. The relative ease in calculating savings and that managers gets "credit" in their cost center for lowering their costs outweighs the project with the harder to determine impact that the manager would have not gotten "credit" for, despite the fact that the underlying data may have indicated the other project had a larger positive impact overall.

Both issues can be handled by showing the full economic impact of a decision at the global level rather than the cost implication at the local level. By sharing information across the organization, it will be easier for managers to know the impact of their decisions and also for their leaders to be able to evaluate them not on local metrics, but on how they have helped to further improve the system.

1.2.4 Learning Experience

The final aspect that needs to be addressed to generate long term success is the way in which learning is created. Novartis managers typically consider a decision by first coming up with an idea of what seems to be worth improving, gathering data, analyzing it and making a conclusion. In addition to the previously discussed challenge of understanding the whole system, there are several other shortcomings with this approach.

First, ideas of what to improve are based on mental models of what already seems important to change. It is entirely plausible that the most important areas are based in areas that are discounted by the organization's current mental model. This problem remains unaddressed when the manager performs the analysis only on the question at hand, missing any potential learnings about the rest of the system.

Secondly, recollecting data to answer each question individually can be intimidating and burdensome, especially for a seasonal product that requires rapid decision making.

Instead of collecting data and answering each question individually, a new approach using a model that possesses all necessary data, but only uses the relevant pieces will transform the way in which analysis is done. Much less time will be spent gathering data and more time can be spent learning from what already exists. Allowing managers to use a system wide model, sometimes called a "management flight simulator", allows an individual to quickly test various scenarios in order to understand how they impact the overall system (Senge 2006, 326). These experiences shorten the amount of time needed for analysis, while also building the intuition necessary for comprehending why the system performs in the ways it does.

1.3 Hypothesis

With the increases in complexity in a multi-stage batch production operation, an approach is needed to guide understanding of the system as a whole. Modeling the financial and manufacturing elements of the value stream will aid managers in their production decision making. As a result, they will better invest and focus resources to improve their organization which will ultimately better serve their customers, patients and shareholders.

1.4 Organization of Thesis

Chapter 1 explains the project motivation, current and desired state analysis and the hypothesis.

Chapter 2 describes the company background and manufacturing process overview.

Chapter 3 discusses the cost collection required for mapping the elements in the production system.

Chapter 4 explores the key parameters in the manufacturing system and model.

Chapter 5 explains the integration of the various components to build the model.

Chapter 6 discusses the various approaches for utilizing the model to serve different information needs.

Chapter 7 summarizes the learnings from the model and provides recommendations for the future.

1.5 Confidentiality

Due to confidentiality reasons, the data provided in this thesis differs from the original Novartis values. Any numbers, timelines or graphs that have been included are either distorted or hypothetical. Also, some decisions are discussed at a high level to illustrate the concepts rather than getting into the details in order to protect confidential plans.

CHAPTER 2: PROJECT CONTEXT

2.1 Company Description

Novartis traces its roots to 1758, when Johann Rudolf Geigy-Gemuseus began selling chemicals and drugs in Basel, Switzerland (Novartis 2010c). The company, eventually known as J.R. Geigy Ltd., merged with Ciba in 1970 to form Ciba-Geigy Ltd. In 1996, Ciba-Geigy merged with Sandoz to form Novartis. The vaccines and diagnostics division of Novartis was formed through the purchase of Chiron in 2006. Novartis Vaccines and Diagnostics currently manufactures over 20 vaccines (Novartis 2008).

For the project to be performed successfully, it must be consistent with Novartis' corporate mission:

> We want to discover, develop and successfully market innovative products to prevent and cure diseases, to ease suffering and to enhance the quality of life.
>
> We also want to provide a shareholder return that reflects outstanding performance and to adequately reward those who invest ideas and work in our company.
>
> (Novartis 2010d)

The project will therefore strive to accomplish both aspects of the mission: making advances for patients and improving financially.

2.2 Product and Customer Description

The product, Fluvirin, is Novartis' influenza vaccine for the U.S. and several other countries. For the 2010-2011 campaign, approximately 40 million doses of the vaccine were delivered to the U.S. (which accounts for the majority of Fluvirin's sales) (Novartis 2010b). This project will analyze all of the production for Fluvirin.

Fluvirin is sold in two different presentation forms: prefilled syringes (PFS) and multi-dose vials. The single use syringes contain .5 mL of the vaccine dosage, whereas the vials contain 5 mL for 10 doses (Novartis 2010a). Due to regulatory differences between the U.S. and the rest of the world, two types of syringes are sold; those will simply be distinguished as U.S. PFS and non-U.S PFS. Preferences by care providers exist for the PFS because of their quickness and ease of use. Others prefer vials because the 10 dose vial is much smaller than 10 syringes and is much easier to store in the often limited space of a refrigerator. Most influenza vaccine manufacturers also sell both the syringes and vials.

Each dose of the vaccine, called a trivalent, protects against three strains of the flu virus. It is produced by blending together three vaccines that target each specific strain, called the monovalents. The three flu strains targeted by the vaccine are H1N1, H3N2 and B.

There are several different types of customers that the product is sold to that then administer the vaccine to patients. Hospitals, governments/public health groups and pharmacies (e.g. Walgreens) are the large groups that deal directly with patients (HIDA, 3). Additionally, vaccines are sold to large distributors that then sell to the small physician offices. Orders are received months in advance and often multi-year contracts are signed. Sometimes, a portion of the agreed upon quantity is allowed to be returned if sufficient patient demand does not exist.

Understanding how the customer values the product is critical in order to deliver to their expectations. The most important aspect of any drug is that it is safe for patients. Unlike other products that can be compared across a spectrum of low to high quality, drug quality is binary, it is either safe or not. Customers and patients both rightly assume that they are receiving safe products from the manufacturers and meeting that expectation is essential. As delivering a safe product is required, the fact that all necessary steps to ensure that safety is a given. Once the safety criteria are met, we can then turn our attention to the other aspects of the product on which the customers and patients place value. (Additionally, longer term innovations, such as new ways of delivering the drug, developing new vaccines or expanding the number of strains vaccinated against, are not the focus of the thesis as they come from R&D. We will only target those attributes which are determined by the value chain.)

Product availability and on time delivery are important to both patients and customers. Patients expect that they will be able to receive a vaccination in the fall so that they do not risk contracting the virus before getting their shot. Unfortunately, during some years in which some strains did not grow well in production, delivery was later than patients expected or did not occur at all. Since the providers want to meet their patient's expectations, they also want the vaccine delivered on time. Not only do they want it early to help their patients, but to also compete effectively against other providers. If a pharmacy does not get its vaccine delivered on time and must turn away patients, those patients may opt to visit a competitor (supplied by a different manufacturer) which does have the vaccine ready. Getting the product delivered early is also a positive signal to the customer that the manufacturing process is running well when they consider increasing their order quantity or signing a long term contract.

2.3 Manufacturing Process

2.3.1 Campaign Cycle

Though the flu season only occurs during part of the year, vaccine production is running almost throughout the entire year in order to provide a sufficient number of doses to patients. Each year, there are actually two separate "flu seasons", one for the northern hemisphere and one for the southern, because of the timing of winter. Fluvirin is produced for both hemispheres, but for the purpose of this analysis, we can focus on the northern hemisphere. Production for the southern hemisphere has a similar timeline, but is shifted and also shorter because of lower demand. Lessons from and the techniques for analyzing the northern hemisphere campaign will be essentially the same for the southern hemisphere.

The production process can be segmented into four main steps: primary, blend, fill/finish and distribution (Novartis). Delivery for the northern hemisphere campaign occurs from July to November. Each stage in the process begins and ends at a different time in order deliver as quickly as possible. Developing what is essentially a new product every year requires certain technological and regulatory barriers to be overcome in order to start each step. Additionally, because the vaccine is different each year, no inventories are held between campaigns and each campaign must therefore start from ground zero.

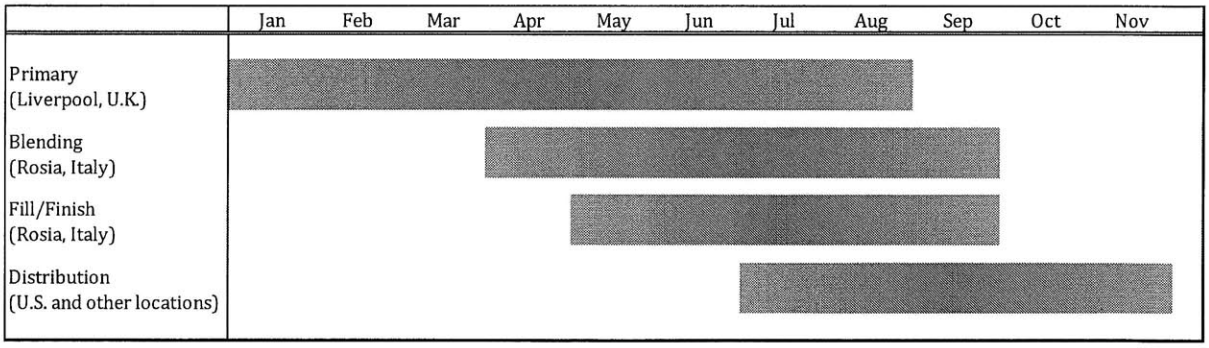

Figure 2: Northern Hemisphere Campaign Timeline

Primary, also known as bulk manufacturing, is the process of creating the monovalent vaccines. Each of the three different monovalents that are produced corresponds to the three strains that are decided upon by the World Health Organization (WHO) (Novartis). Though the official announcement for the strain selection is in February, production may begin as early as January for the strains that are anticipated to be recommended. Primary manufacturing takes place on a fully dedicated production line in Liverpool, United Kingdom. Inventories of the bulk monovalent are built up until the blending process can begin.

Within a few months, the blending process is ready to begin. This team takes the monovalent from each of the three strains and blends it together at the proper ratio to make the trivalent vaccine.

Each batch of the trivalent vaccine is run through a fill and finish process line to fill the fluid into the appropriate presentation (vial or syringe) in order create the final product. Fill and finish as well as blending (together commonly referred to as "secondary") are located in Rosia, Italy. In addition to the various quality tests that have been performed throughout the process, a final set of regulatory approvals is needed before the product becomes available for use. Pallets of syringes and vials are then shipped in climate controlled containers to the customers.

As the product moves through the manufacturing process, different financial transactions take place for legal and accountability reasons. Each government authority expects to collect an appropriate amount of taxes for the work being done. The production process is owned by the Liverpool site. When the bulk material is sent to Rosia, the Italy site charges Liverpool for materials, allocations and a markup. These additional costs are expected by the tax agencies. Then, the Liverpool site "sells" the final product to the Novartis group in the country of the final sale (usually in the U.S.) through a transfer price. This group then sells the product to the customers. Though these transactions need to occur for legal reasons, we will need to consider what numbers are appropriate to consider in our analysis.

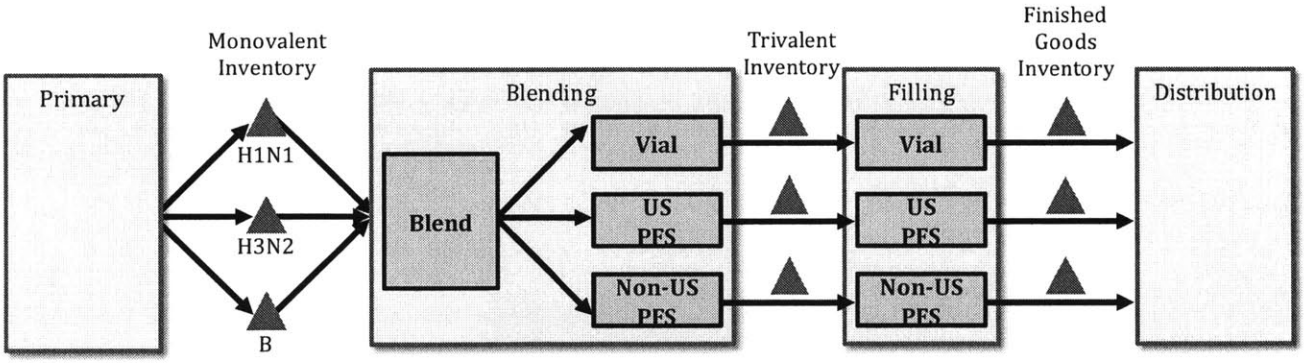

Figure 3: Multi-Stage Batch Production Process

2.3.2 Daily Production

Within the primary production process, there are three main steps: primary incubation unit (PIU), upstream and downstream. Though it takes several days for a batch to be completed, the process is pipelined so that one batch starts and another finishes each day (a.k.a. "batch a day"). Setting a fixed cycle time of 24 hours helps standardize and simplify work for the employees so that the organization benefits from "economies of repetition" (Jones 2006, 29).

Only one production line exists in primary, so the schedule alternates in order to make the three different monovalents. Figure 4 shows batches scheduled in each stage of the primary process. Some stages require multiple days (e.g. first incubation in PIU) so multiple batches would be present at those stages.

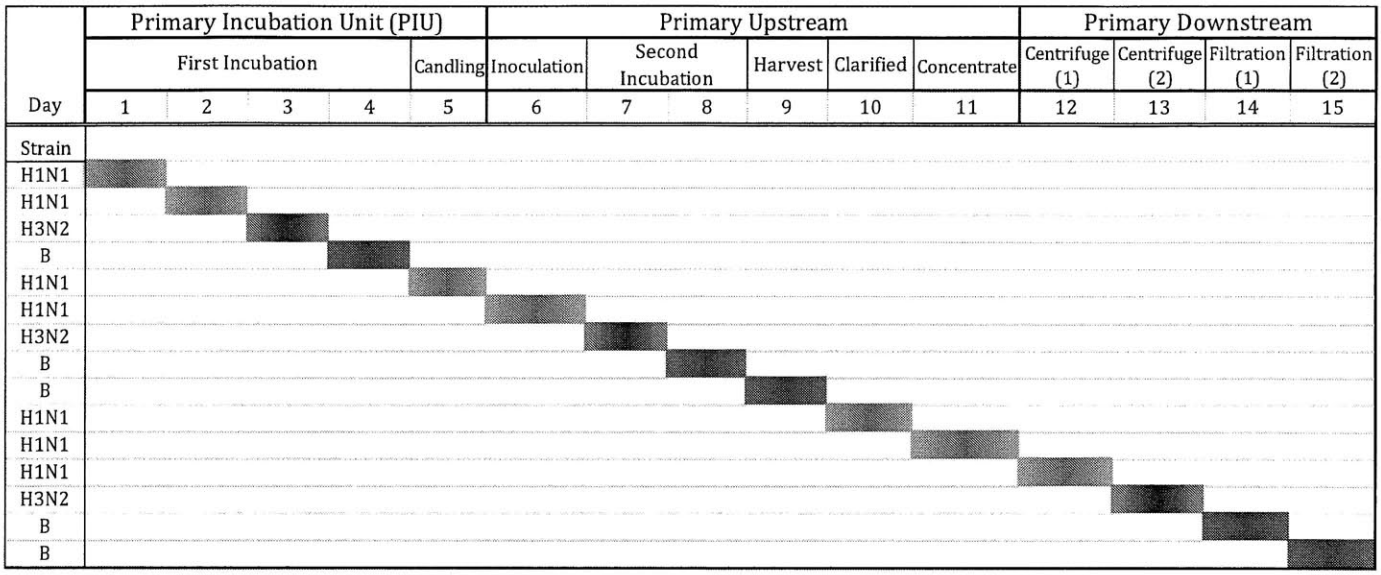

Figure 4: Batch Scheduling in Primary Process

The start of a new batch begins with the delivery of special chicken eggs to PIU. For several days, they are housed in large climate controlled incubators for the eggs to mature. When ready, the eggs go through a candling process used to screen for bad eggs which cannot be used for production and could also later lead to contamination. Due to regulatory process standards, only a certain number of good eggs can be sent to upstream. Because of the inherent variability in the number of bad eggs, extra eggs are always purchased in order to always ensure sufficient supply to the next step. Unfortunately, any time there are more eggs than needed, they must be discarded and cannot be used in the future.

Eggs travel from PIU into the upstream area. Each egg is injected with the seed virus (inoculation), and the virus will then grow inside the egg during a second incubation period for the next several days. The type of seed that is used will determine which flu strain the batch will produce. Once complete, the eggs are opened (called "harvesting") and the allantoic fluid (the fluid inside the egg that now contains the virus) is removed. The allantoic fluid is clarified, concentrated and inactivated (i.e. the virus is killed).

Further processing takes place in the downstream area. First, the fluid is processed through zonal centrifugation which further purifies and concentrates it. A second centrifugation step separates the

hemagglutinin (HA) and neuraminidase (NA) surface antigens from the virus. (The HA antigen is what the human body recognizes in the vaccine in order to develop an immunity.) After a final filtration step, the result is the monovalent antigen containing the HA.

Each day results in a new batch and each batch produces the same quantity of monovalent fluid. However, due to the differences between strains, losses that occasionally occur and the natural variability in a biological process, the potency (the number of doses that can be made) can vary substantially (by tenfold). A sample is taken from each completed batch in order to analyze its potency by means of a serial diffusion test (SRD). To eventually make a trivalent that has the right amount of each monovalent, the schedule is created to balance the number of available doses. A strain that often has a high potency will be scheduled less frequently that a strain that averages a low one. Figure 4 shows how the low yielding strain, H1N1, would be scheduled more frequently than the higher yielding strain, H3N2.

When sufficient inventories exist of all three monovalents and all quality tests are complete, the blending process is ready to begin. In order to blend the proper ratio of the monovalents, two pieces of data are necessary. First, the potency results of the SRD tests determine how much HA is in each batch. Second, the blend target, the amount of HA going into each dose, must be determined. The final vaccine is required to have 15 ug of HA of each of the three strains. However, the same amount of HA in each dose is not the same. This 15 ug requirement is a minimum during the entire shelf life of the vaccine. Over time, the amount of HA in a syringe or vial actually declines. Therefore, in order to have 15 ug of HA by the end of the shelf life, more than 15 ug needs to be filled. Some strains of the vaccine are very stable, with little decline over time, so only slightly more than 15 ug is needed. Other strains however, are unstable and significantly more than the minimum is needed to compensate for the eventual decline. Once the amount of HA is known in each batch and how much HA of each strain is needed, the appropriate volume from each monovalent batch can be combined to make a trivalent batch with the desired number of doses. Phosphate buffered saline (PBS) is added to bring the fluid to the desired volume. (A slightly different chemical process is used to make the trivalent for the different types of presentations, so it must be known in advance which final form is needed.)

Once the trivalent is complete, the large volume of fluid is ready to be filled into one of the two presentations. In order to accommodate the statistical fluctuations in the filling process, a target fill level is set above the .5 mL dosage level (overfill) so that any presentation that receives a quantity below the target will still be above the required amount. Each filled vessel is labeled, boxed, tagged with a leaflet and shipped. When all final regulatory tests are complete for that batch, it is released for use.

2.3.3 *Manufacturing Challenges*

One of the complicating aspects of the flu vaccine manufacturing process is the inherent uncertainty in both the supply and demand. On the supply side, though some initial data might exist, it is often not fully known how many doses a batch will produce on average for each strain. Lab tests and historical data are typically considered the best indicators of future values, but can often be significantly off due to the challenge in predicting the performance of a biological process. Depending on the particular year, it is not uncommon for the average number of doses in a batch for one strain to be 2-4 times higher than another strain. Additionally, not only is it difficult to forecast the yield for a particular strain, sometimes which strain will be selected is unknown. Novartis and the other flu vaccine manufacturers work with the WHO throughout the year to track the rise of the different strains. However, sometimes when making a decision well in advance of the start of a campaign, it is not always clear which strains in the world will be the leading strain by the time of the decision.

Coping with demand is also a challenge due to uncertainty. First, influenza outbreaks are difficult to predict, and because some people get a vaccine in response to the severity of the flu season, it can be hard to know exactly what patient demand will be (Danzon, Pereira, and Tejwani 2005, 714). Secondly, uncertainty in a seasonal product is especially challenging because most production will occur before the actual flu season. This uncertainty is mitigated somewhat with the contracts that the company has in place with its customers which help to stabilize and reduce the level of uncertainty. Finally, uncertainty also exists in the marketplace, where the actions of competitors can have a significant impact. During the 2010-2011 campaign, flu vaccine manufacturer CSL, which had accounted for several million doses of the flu vaccine sold in the U.S. in the previous year, had quality concerns and completely dropped out of the market for the entire year (Sands 2010). Significant decreases in production or late deliveries leave potential gaps in patient coverage as well as opportunities for other manufacturers to step in.

Compounding the challenges associated with dealing with the uncertainty are the long-lead times associated with the production system. Regulatory approvals needed for production system changes, significant training time for employees and slow biological processes (e.g. chicken raising for eggs production) all reduce the system's ability to quickly respond to changes. While it would be good to reduce these times to help develop a more agile system, it is sometimes not feasible. We will need to make sure that any analysis performed understands the inherent uncertainty that is faced and we will consider efforts that will help smooth flow.

2.4 Project Scope

Based on the company's mission, we have already determined that we want to both improve on delivery to patients and increase its financial performance by seeking a global optimum. Pursuing a global optimum by analyzing all of Novartis is clearly impractical, so we must find a way to achieve that objective in a practical way. Defining a sensible set of system boundaries will enable our analysis at a reasonable level while still being consistent with global objectives. The boundaries will exist around the appropriate sites, product lines and across time.

First, in order to identify a global optimization for the company, the boundaries of the value stream extend from the beginning at the Liverpool manufacturing site, through Rosia to the U.S. distribution to the customer. End to end value chain evaluation will ensure that each stage in the process understands its importance to the entire system and no local optimization occurs based on internal money transfers. For example, if we were to only focus on the Liverpool site, making a change that increases the sale of an additional dose would add the value to the site of the transfer price rather than the actual commercial value. Focusing just on the site would undervalue this transaction and distort the perception of what really should be done.

Secondly, cost mapping will be exclusively for the Fluvirin products as decisions will be treated to only affect those. Fluvirin is Novartis' only flu vaccine in the markets that it serves, so no cannibalization would occur to its other products if more Fluvirin is sold. Additionally, we will assume independence between Fluvirin and any other vaccines in Novartis' portfolio because of the different customer relationships that exist for a seasonal vaccine and the other vaccines.

The third boundary is across time and will be set for the one year campaign. This is the natural boundary for most planning discussions that take place because it captures the full production and sales cycle. Additionally, many of the contracts and period costs are committed for the full cycle. Finally, most of the managerial questions under consideration are for a single campaign because they are usually consistent with what also should be done long-term. To perform any analysis that needs to be evaluated over several years, we could look at one year at a time in order to use tools like NPV.

CHAPTER 3: DATA COLLECTION OF COSTS

3.1 Need for Cost Analysis

Financial mapping across the value stream is one of the key tools necessary for understanding the financial impact of a decision in this type of environment. As with every managerial accounting decision, we must continuously seek to identify the revenues and costs that are relevant for the discussion. To aid with the relevancy question, we can rely on the three previously identified system boundaries: sites, products and time. We will examine Fluvirin financials over the course of a campaign in Liverpool, Rosia and during distribution.

This project aims to develop a tool used for improving decision making in different scenarios (many of which have not even been defined yet by management). If it is crucial to use just the relevant financial data for each decision, how is it possible to collect cost information that is relevant for various situations? Instead of first narrowing the data down to be relevant for a specific situation, we will gather the costs appropriate within the defined boundaries of the potential questions and then utilize an approach that for each particular situation will ignore the irrelevant data. This will help reduce the amount of data that each person will need to collect for their question in the future because most of the relevant data will already have been gathered. Gathering extra data up front saves the members of the organization from having to duplicate data collection efforts.

We will therefore need to collect all costs that would be changing in response to a production system change (i.e. marginal or variable costs, also called totally variable costs in throughput accounting). Costs that are for running the site would not be directly changing with the decision are operating expenses. The details of these costs play a secondary role in the decisions because we do not need the specifics in how they would change, only by how much. For example, increasing capacity by hiring additional employees in a particular group would be an example of an increased operating expense that would need to be calculated based on the specifics of the project by the user. Based on our generalized methodology, we will provide the user with the information of how the rest of the system would respond and they will combine it with the change in the operating expense specific to their situation to inform the decision. One-off expenses will be treated in the same way.

Based on historical practice and interviews about future plans, it is expected that very few of the managerial decisions that we are addressing would impact operating expenses. This is consistent with findings from throughput accounting research that operating expenses are relatively insensitive to production changes (Noreen, Smith, and Mackey 1995, xxvii).

3.2 Categorization of Costs

Finding the cost data for a vaccine, or any manufactured good in a similar production system proves to be more nuanced than would originally appear. Relying on the fully allocated standard cost calculation, as has previously been discussed, would not be useful for the types of discussions in question because they include fixed costs. However, simply identifying the direct variable costs in order to calculate a single value for the variable cost per dose is not sufficient.

What makes the financial structure interesting in this instance is why it is critical to split the costs across two dimensions rather than aggregating all the costs into a single value. Multi stage batch production under one setting can have a significantly different cost structure than when the process is changed. In many settings, it is reasonable to analyze a product's cost as the summation of all the individual unit costs. For example, the marginal cost of producing an additional car can be approximated by the sum of the marginal cost of all the additional parts like the wheels, doors and engine. Cost structures like this exist because each piece is just a unit marginal cost. Fluvirin production is different; it cannot be categorized as the sum of unit costs because it is produced in batches that can change. As the production system changes, the entire cost structure is liable to change in various ways. Therefore, we need to differentiate the costs to handle those changes by breaking costs up along the value stream and by cost type. Splitting costs along the stream will allow the costs to move independently in each stage of the process. Differentiating costs as unit or batch will handle changes in production volumes appropriately. A couple of very simple examples representative of how the Fluvirin manufacturing changes might impact costs demonstrate the need for segmenting along these two dimensions.

A two stage sequential process that currently requires 10 batches in the first step and then 8 in the subsequent process has a particular cost structure. If efficiency improvements are made in the first stage so that now only nine batches are needed to produce the same number of units, *some* variable costs are reduced (savings are recognized by having one fewer batch in the first stage, but no savings are made in the second stage). Without separating the process steps and only using a single cost to characterize the process, it would not have been possible to see that one batch cost was saved for the first processing step.

Similarly, batch and unit costs need to be distinguished. Marginal unit costs are those costs that change with respect to the number of units being produced (like the car example). Batch costs are only those costs that change with respect to the number of batches being produced.

- (Marginal) Batch Cost: Cost of producing one more batch
- (Marginal) Unit Cost: Cost of producing one more unit within a batch

(It is important to note that products that are made in a batching environment do not necessarily have significant batch costs.)

Consider another simple example, this time with a single manufacturing step, but that has both unit and batch costs. A process may currently require five batches to produce the desired number of items. With an efficiency improvement that reduces the number of necessary batches to four to produce the desired number of items, *some* variable cost savings would be witnessed. Because the number of units stays the same in both situations, the unit variable costs would stay exactly the same, only one batch cost would be saved. Again, trying to characterize a product by using a single value for its costs would have led to problems when analyzing a change like this.

3.3 Approach

Once it is known how the different variable costs need to be categorized, a process is needed to collect the data. Getting variable costs seems easy given the elaborate IT systems that are such an integral part of today's companies. Unfortunately, the challenge is transforming the sea of data into the useful information (Goldratt 1990, 4). The challenge exists primarily because these IT systems are designed to be broad enough to handle a generic manufacturing system and are intended primarily for financial accounting. Rather than creating reports for how much each group is spending, a managerial accounting system needs to identify the right data to inform the right decision.

Costs are tracked in the accounting system in two different ways: in the bill of materials (BOM) and in a cost center. The general concept behind this traditional structure is that the variable costs are part of the BOM, while fixed costs are included in the cost centers. Unfortunately, the lines between the two types of costs are not as black and white as this, thus requiring additional digging.

Materials on the BOM appear either in the raw or semi-finished state. Raw materials for Fluvirin are straightforward variable costs as they are simply calculated as the price paid to the supplier[1]. All materials (except for eggs, which will be discussed) in Fluvirin maintain a steady value and have a long enough shelf life such that it is appropriate to value them at the price paid. Semi-finished goods are processed materials that have already had fixed costs allocated to them. When gathering the costs for items that include semi-finished goods, only the variable costs should be included.

---

[1] Using the price paid is not always appropriate in every situation. The relevant cost is really the value of the material at the time of the decision. Materials that appreciate in value (wines, antiques, etc.) or depreciate (perishables, electronics, etc.) should be evaluated at their new value.

Analyzing the cost centers is a bit more complicated because they serve as a catch-all for everything that is not included in the BOM. While it is clear that those materials in the BOM are variable costs, what was initially unclear to management was how many variable costs are not included in the BOM. In some manufacturing settings, those variable costs not included in the BOM are negligible. But in biopharmaceuticals where several costs like quality testing, hazardous waste disposal, cleaning and maintenance are variable, but do not directly go into the product and are therefore not in the BOM, additional effort must be made to calculate those variable costs. Ignoring these costs would lead to an underestimate in the true marginal cost of production and distort the financial realities of a decision. Since no analysis had previously been performed on the magnitude of these indirect variable costs, there was significant concern that it could be substantial. To identify them, these costs need to be distinguished from the fixed costs that are also included in the cost centers.

To ensure the accuracy of the cost data for the cost centers, we can collect it in a bidirectional way and verify the results. We can approach the data from the top down by interviewing the various managers who own the different segments. Then, as a bottom up exercise, data can be collected from the financial database. Comparing the cost estimates from the two approaches will confirm the validity of the costs or highlight where additional effort needs to be focused to improve the accuracy of the numbers.

In summary, the direct costs will be calculated from the BOM, whereas the indirect costs will be extracted from the cost center data.

3.4 Direct Materials Costs (BOM)

Collecting the cost data for the direct materials is the most straightforward as it can be easily pulled from the BOM. As the BOM is constantly kept up to date with the precise quantities needed for each production step, the data can be trusted as highly accurate.

Mapping and categorizing the different costs is the next step. Within the ERP system, each distinct manufacturing step requires a separate process order (and therefore a separate BOM). In the Fluvirin process, different process orders are required for primary production, blend and fill/finish. This breakdown of process steps is consistent with how we wish to perform our analysis making it easy to map those costs. What creates confusion with users of the ERP system is that rather than just keeping each process cost separate, it attempts to accrue all of the previous costs. Not only does this create issues when understanding how costs change with batch changes as with the earlier example, but it uses different fractions of batches to be consistent through each step. Additionally, fixed cost overheads are allocated through each step of the process which distorts the real cost of each step. Again, we can just use the

variable costs from each step in the BOM and ignore all the additional complexity of how the ERP system is used.

Once the costs are mapped, the final stage is to categorize them as unit or batch costs. The managers already have a good understanding of what is appropriate to consider for each group. In Fluvirin, there are very few direct materials costs associated with the number of units being produced. Those unit costs predominantly come in at the end of the manufacturing process in the fill/finish stage. To make an additional dose of the vaccine (i.e. the marginal unit cost), the only additional material is a vial or syringe and the corresponding packaging.

Because of the nature of vaccine production, all other direct materials can be considered batch costs. This is a hugely important fact in understanding the economics of the manufacturing process and what distinguishes it from many other systems. Through many stages of the manufacturing system, Fluvirin has an elastic batch size: the size of the batch can change without impacting the rest of the system. Most systems have an inelastic batch size: the batch size cannot change without making other cost changes in the system. A good example of an inelastic batch process is transportation in a truck which has a fixed batch size (maximum volume or weight). One could not simply add additional packages into a full truck, a second truck would be needed or one would need to get a larger truck (presumably at an added cost). Elastic batch sizes are different because they can increase in size without an added cost or an increase in processing and quality testing time (within a reasonable range).

The Fluvirin primary process is elastic because it is not defined by physical parameters like volume or weight, but instead by potency. Changing the potency (by either increasing or decreasing it) would have no affect on the actual batch costs for a particular stage. A 5 liter bag costs the same regardless of whether it has the equivalent of 100K doses or 400K doses. The manufacturing process standards are designed around volumes and times (e.g. filter the 100L for 1 hour or concentrate 600L down to 200L over 3 hours). Changing the potency would therefore not affect the batch cost or even the processing time. The rest of the direct materials costs fall into this batch cost category that is elastic, some examples include:

- Bags: Used to store the fluid at each intermediate step
- Rigs: Joints and tubing used to transfer the different fluids
- Filters: Removes impurities in fluids
- PBS: Buffer solution used to dilute fluid for different steps
- Etc…

Eggs represent the most significant batch cost and require some additional discussion. Unlike the other direct materials, the assessment of costs for the eggs is different. All of the other materials have sufficiently long shelf lives that they can be used in future campaigns if some are not needed during the year. Eggs however are perishable and become a sunk cost once they are delivered. For very short term decisions, the cost of any eggs in inventory should be ignored as they are sunk. Similarly, during certain periods of the year, some egg quantities are already agreed upon in a contract making them a committed cost and have effectively no resale value. Other eggs that are bought through options or on the open market would simply be tracked at their cost. Differentiating between the different types of eggs purchases will be critical to properly assessing changes in production volumes so that we accurately capture the economic realities of a plan under consideration.

3.5 Indirect Costs

3.5.1 Indirect Materials, Consumables, Quality

Identifying the indirect material costs are more challenging than the direct materials because they are not precisely defined in the BOM. Instead, materials used outside the BOM are just added into a cost center and have significantly less traceability than materials that are scheduled through a process order. Nonetheless, interviews with each front line manager of the different cost centers highlight the materials that are used during a typical batch. Additionally, filtering through the ERP system is possible. Material costs are assigned to the cost center when they are debited from inventory, not necessarily when they are used, making it hard to know precisely the usage levels. By searching over an extended period of time it becomes possible to distinguish the materials that vary with production and ignore the fixed costs of the materials that are used independent of production. The types of items that exist in this category include quality testing supplies and clean-in-place (CIP) chemicals. Estimates provided by the cost center owners proved to be consistent with the data analysis and could therefore be used reliably. Despite the fact that each individual cost center owner knew the details of their use, this is the first time effort had been taken to pull together all of these indirect materials costs and distinguish between fixed and variable.

3.5.2 Equipment Use and Repairs

Understanding the costs associated with production extend beyond simply materials and into equipment usage. Again, in order to inform the key types of decision we are interested in, only the marginal costs associated with producing an additional unit or batch should be considered. Traditional cost accounting considers the number of equipment hours used in performing a task to assign a labor cost and a cost of using the equipment (i.e. the depreciation). Similarly, the labor cost of having a repair staff

is allocated to the process. All of these items are committed costs and are irrelevant for how the system would be impacted by what we are concerned with. The only cost components that are relevant are the repair costs that actually vary with production volumes. Because of the nature of the vaccine production process, there are actually several highly expensive machinery components within the equipment that need to be replaced for quality reasons after a certain number of batches has been produced (often after every batch). These costs would be affected by our changes and need to be included in the cost analysis.

3.5.3 Utilities and Waste

One of the large total costs for Fluvirin is for electricity, natural gas and waste, but little is known about how much of those costs are fixed or variable and therefore it is impossible to know how those costs would change with our decisions.

First, we can analyze the electricity usage to understand the variable cost of the utility during production. Comparing the electricity levels when a batch is being produced to a baseline when no production is occurring helps gauge the impact of production. Figure 5 shows that for two similar periods of time (so that other potential changes in electricity are negligible) there is very little impact on usage by production and it is therefore essentially a completely fixed cost. Despite the large pieces of equipment used during parts of the processing of the vaccine, the reason electricity is almost all a fixed cost is because of the heating ventilation and air-conditioning systems that are always running for the clean rooms.

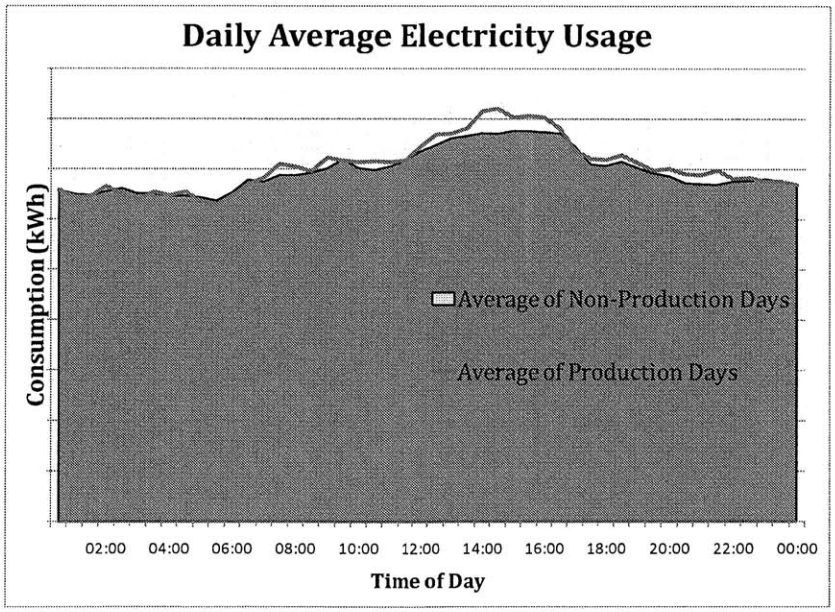

Figure 5: Electricity Usage During Production

Similarly, some pieces of equipment use natural gas. However, unlike the electric equipment, much of the gas powered machinery is only active during manufacturing and so there is a much larger portion of the cost that is variable. Comparing similar days when production is active to the days when it is inactive isolates the amount used for production. Figure 6 shows that approximately a third of the natural gas costs are variable for the primary process step.

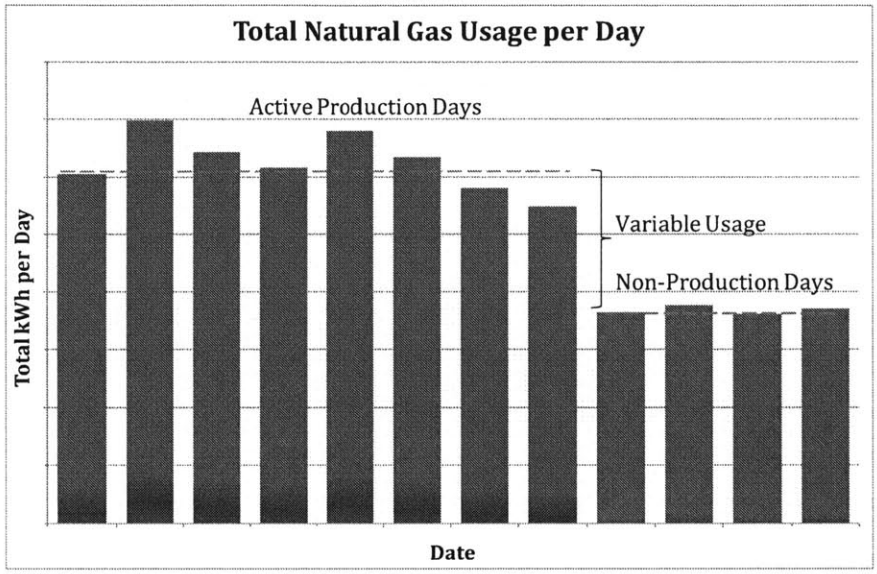

Figure 6: Gas Usage During Production

Waste costs are quite high in vaccine production because of the disposal requirements for the egg waste and hazardous chemicals. It is predominantly a variable cost because the expensive fees for special disposal vary directly with the production volumes (fixed waste quantities are inexpensive because it is simply general waste). Analyzing the waste removal contracts and historic charges in comparison to the volumes identifies how much is being spent in each production area for each batch produced.

3.5.4 Logistics

Freight charges are also a significant variable unit cost and are actually not normally part of what is considered in the "product cost" in the traditional accounting system (the cost is charged to the site rather than per unit). Because we are considering the total cost through the value stream, all logistics costs are considered, rather than those for a particular site. Most of the logistics costs are in the final stages of the process because of the special handling requirements of shipping in a climate controlled container and because many of the raw material costs already have the inbound shipping charges included.

3.6 Cost Summary

Understanding the breakdown of variable costs along the value stream and categorizing them as unit or batch is an essential step to inform our production decisions. This breakdown will eventually allow us to see how the cost structure of the product changes and the manufacturing process changes. Too often costs are just lumped together to be treated as all the same, but understanding the differences is critical. We have discussed how standard costs in traditional accounting overestimate the real cost of manufacturing through allocations and would therefore not be appropriate to use for the types of changes being proposed. However, simply using the direct material costs from the BOM would also be an inappropriate estimate for the total variable cost for many products including Fluvirin. Using just the direct material cost would (potentially significantly) underestimate the true variable costs distorting our analysis. The costly indirect materials, special replacement parts needed for each batch, variable waste costs for chemical processing and freight costs comprise a substantial portion of the total variable costs in vaccine production and cannot be ignored.

While the details are confidential, an example of how significant these indirect costs can be in this manufacturing setting is in the primary production step in Figure 7. (Each manufacturing stage has a different breakdown of total variable costs. For example, logistics costs in primary production are low,

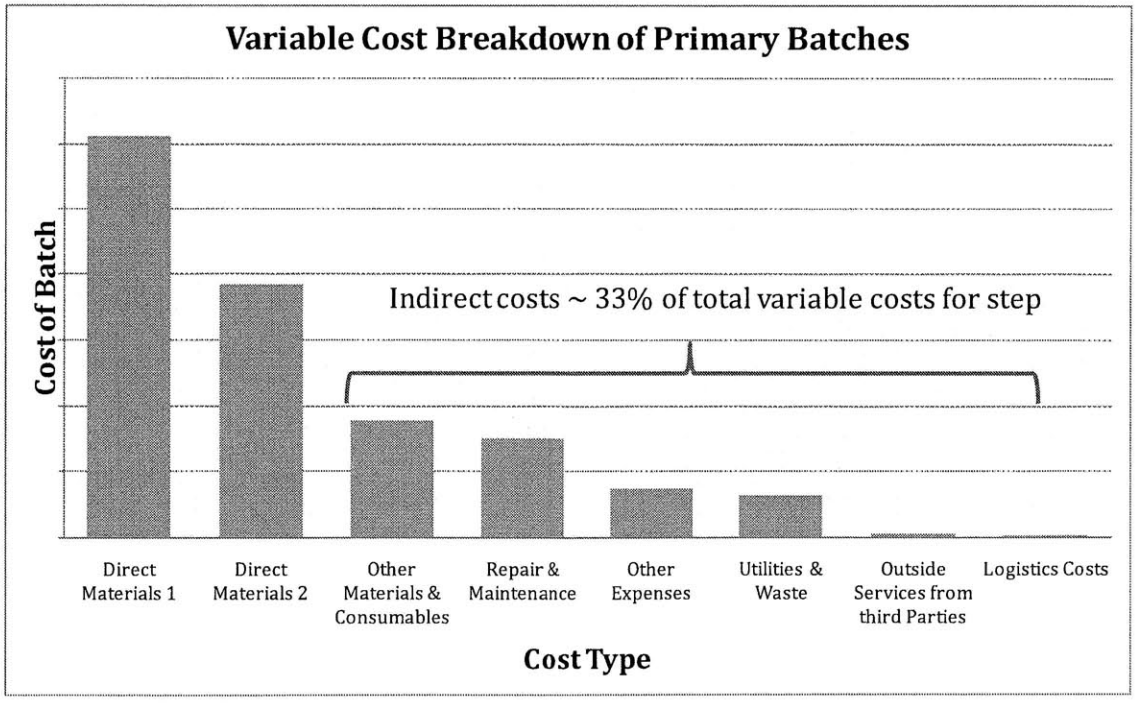

Figure 7: Variable Costs in Primary Production Stage

whereas they are the second highest cost in fill/finish because of the final delivery costs.) Approximately 33% of the total cost in primary production is not a direct material. This final insight has two important applications. First, in a manufacturing environment with extensive indirect work (e.g. quality testing) and expensive special processing (e.g. cleaning and waste disposal) this type of analysis is important rather than just using direct materials for the variable cost because of its significance. Secondly, these numbers show senior management how much the real marginal cost is compared to the standard cost, because despite being much higher than the direct materials cost, it is still significantly lower than the fully allocated standard cost. Historical uses of standard costs seem to anchor a high cost value in the minds of the managers, so even though they know that standard costs includes fixed cost allocations and should not be used as an estimate for a variable cost, there are no other number to compare to. Even without doing any additional special analysis, now knowing the real variable cost instead of a standard cost is tremendously valuable because it reshapes managerial intuition.

Though the details are confidential, the general breakdown of variable costs identifies primary production as the main driver followed by fill/finish and lastly blending. Primary and blending are almost exclusively batch costs, whereas filling is nearly all unit costs.

Once an exhaustive data collection has been performed on what the variable costs are, we can then move forward with the faith that our future analysis will accurately capture the true economic realities of the manufacturing system.

CHAPTER 4: MANUFACTURING PROCESS DATA

Understanding the cost breakdown of the product is only half the necessary data needed for understanding the system. The other pieces of data are the characteristics of the manufacturing process. Traditional cost accounting and other allocation approaches both focus attention on the time (or activity) used at each processing step for allocation purposes. We know however, that many of these are fixed operating expenses and therefore allocations would not be relevant for our analysis. As a result, the allocations used in these processes would distort the economic realities of what would be changed due to a managerial decision.

Throughput accounting instead analyzes the system as a whole by using process times to identify the constraint and then use that information to drive the analysis. Goldratt, Corbett and others have provided several examples of how to evaluate the impact of a decision by playing out the new scenario using throughput accounting techniques (Goldratt 1990, 67; Corbett 1998, 11). Those approaches utilize information about the cycle times for each processing stage to calculate the number of units produced and financial throughput. When a change is proposed, the calculations are updated to reflect the new value (e.g. reducing a cycle time or changing a process routing) and then compared to the original situation to understand the overall impact.

Analyzing the Fluvirin process builds on the concepts of throughput accounting, but differs in a few key areas. First, the relationship between cycle time and throughput are different in the batch a day process. Normally, a shorter cycle time means that more units can be produced in the same period of time. However in batch a day, if the cycle time reduces, the remaining time in the day sits idle and no additional units are produced. Instead, throughput in a batch a day process changes by changing the batch size. Therefore, instead of focusing on cycle times, the Fluvirin analysis will evaluate batch sizes.

Focusing simply on batch sizes still leaves an important gap in the process parameters that managers actually control. Specifically in primary, Novartis engineers do not directly modify the batch size; instead, they change the underlying parameters that would then lead to a different batch size. Providing them with the key process parameters that control batch size, rather than the batch size itself is far more valuable to connect their work to the overall system.

Finally, throughput accounting has tended to focus on production in which there are no batch costs and as a result, treat the totally variable cost as a single value. As was discussed in Chapter 3, Fluvirin's multi-stage batch production requires costs to be distinguished between unit and batch as well as mapped to their respective processing stage. All of the necessary cost data was collected in Chapter 3,

the only additional effort necessary is to ensure that the stages used in the process data match up to the financial stages (primary, blending and filling will again be used as the three key distinct stages in the vaccine process).

4.1 Calculating Throughput

4.1.1 Primary Incubation Unit (PIU)

Determining the egg throughput for PIU is quite straightforward. It has more capacity to process eggs than primary upstream has to receive eggs. The value is set both by the physical capacity that primary can receive as well as by the current regulatory approval.

4.1.2 Primary Manufacturing Throughput

The quantity that is the output of primary is not a discrete number of units or a volume of fluid (the same volume is produced each day), rather it is the potency which is measured in the amount of HA antigen. A few main variables determine the amount of HA that would be expected to be produced in a batch. HA is refined through the primary process and originates in the allantoic fluid that is harvested from each harvested egg. The number of harvested eggs is simply the number of total eggs going into primary multiplied by the percent of those eggs which are harvestable (i.e. good). Despite some initial screening in PIU, not all eggs are at the quality level appropriate for use. For example, harvesting dead eggs would increase the bioburden level and potentially create an unsafe product. Combining these variables results in the following equation:

$$Amount\ of\ HA\ per\ Primary\ Batch = Egg\ Quantity * Harvest\ Yield * HA\ per\ Harvested\ Egg$$

Egg Quantity		300,000
Harvest Yield		90%
HA per Egg	H1N1	20 ug
	H3N2	29 ug
	B	24 ug

Figure 8: Primary Parameters

(HA per harvested egg is also commonly referred to as the strain yield.)

It is important to note that the amount of HA per (harvested) egg is different between each of the three strains. This is why the resulting batch sizes between each strain differs. Each seed virus grows at a different rate in the eggs, so when they reach the appropriate time to be harvested, there is a different amount of virus in each egg and thus a different amount of HA.

Each of the parameters can actually be changed over time. Creating additional capacity to process more eggs or improve the quality of eggs coming in would increase the total number of good eggs. Also, developing a better growing seed would lead to more HA per egg and thus larger primary batch sizes. Finally, the variable "HA per (harvested) egg" is not the total amount of HA that exists in each egg, but rather the amount that is "captured" from each egg from the processing techniques. More HA actually exists than can be captured. Improving existing techniques or developing new processes can lead to capturing more HA. Until primary batch sizes are understood in the context of the rest of the processes and mapped with the appropriate financial values, it is not possible to know which are worth the cost and effort (some would require regulatory approval).

To simplify the later analysis, it is helpful to convert all throughputs into the number of doses of the trivalent that can be made. Translating the amount to HA from primary into its corresponding number of doses requires knowledge of how the monovalent is going to be blended (i.e. blend targets) and how much will be filled.

4.1.3 Blend Targets and Overfill

As was mentioned in Chapter 2, the monovalents are blended at different ratios in order to ensure that a sufficient amount exists throughout the life of the vaccine. These blend targets are set by Novartis and the regulatory agencies based on the stability data. Also, blending takes into account the potency of each monovalent so that the same number of doses would be present for each strain in the trivalent.

$$Trivalent\ Volume = \frac{Amount\ of\ HA}{Blend\ Target}$$

Fluvirin uses a consistent volume for each trivalent batch to be blended containing a specific number of doses. Since the primary batch potencies are always different, primary batches are combined or split to get the numbers correct. After the three strains are combined, PBS is added to buffer the trivalent to the appropriate volume.

Additionally, the overfill percent is specified by the filling team as to how much each presentation (syringe or vial) will be overfilled in order to ensure dosage volume is met during any fluctuations in fill quantity.

$$Potential\ number\ of\ doses = \frac{Trivalent\ Volume}{Fill\ Quantity * Overfill}$$

Blend Target	H1N1	36 ug/mL
	H3N2	34 ug/mL
	B	32 ug/mL
Fill Quantity		.5 mL/dose
Overfill		110%

Figure 9: Blend and Fill Parameters

When all of the equations are combined, the different variables that engineers control throughout the entire process are mapped to the primary batch size for a batch that was properly made (losses are discussed in a later section).

$$Primary\ Batch\ Size = \frac{Egg\ Quantity * Harvest\ Yield * HA\ per\ Egg}{Blend\ Target * Fill\ Quantity * Overfill}$$

Again, because some of the variables differ for each strain and because primary is only one step in the process, it is not immediately clear how a change in a parameter, such as HA per egg, would affect production, let alone financial performance. Chapter 5 will discuss how those changes can be understood through the use of the model.

4.1.4 Secondary Manufacturing Throughput

Determining the throughput for both blending and filling is much more straightforward. Unlike primary which has a batch size that changes, batches are created in blending and filling to have a fixed quantity.

Blend Batch Size		1,000,000 doses
Fill Batch Size	Vials	1,000,000 doses
	PFS U.S.	250,000 doses
	PFS non-U.S.	250,000 doses

Figure 10: Batch Sizes for Blending and Filling

4.2 Process and Failure Losses

During the manufacturing process, two types of losses occur which erode the number of doses that are actually produced. First, "process losses" happen during the normal course of operations and are inherent to the existing manufacturing process. Second, "failure losses" arise as a result of a problem in which the intended procedures were not fully followed. While many of the losses are already measured, the understanding of how a loss in a particular area affects the overall system is not known. Identifying and quantifying as many of these losses as possible will ultimately increase the accuracy of the analysis and help to show their impact on the system. This will then allow managers to better prioritize their projects in a way that seeks a global optimum.

Process losses are an intrinsic aspect of the current methods. For example, when the trivalent is used to fill the syringes or vials, a small percent of the vaccine is wasted because a residual amount is left in the tubing of the filling machinery during changeovers. These losses are calculated by comparing the theoretical number of doses that should be produced in one step to the actual number of doses produced in that step for a perfect run (a batch that did not have any failure losses). Measuring process losses within primary manufacturing is not possible because the theoretical number of doses is unknown since the only measure of potency happens at the end of primary. Improvements in primary will be represented by an increase in the amount of HA per egg captured, rather than lowering the process losses value.

Failure losses occur when the process is unable to be followed and material is lost as a result. For instance, if one of the egg harvesting machines is temporarily down, any eggs in the batch that are not processed within the specified window of time must be discarded for safety reasons. (In most manufactured goods this type of loss would never exist. Because most stages are decoupled, an inventory would simply accumulate in front of the downed machine until it was operational.) Due to the nature of vaccine production, specifically the use of perishable materials in a long fully-coupled process that is manually intensive, primary manufacturing can sometimes have significant failure losses. Sometimes an entire batch can be lost. Computing an expected failure loss is based on the total number of actual doses produced during the period compared to the potential number of doses that would have been produced had there been no failures.

		Process Losses	Failure Losses	Total Losses
Primary		N/A	5.00%	5.00%
Blending		0.20%	0.50%	0.70%
Filling	Vials	4.00%	1.00%	4.96%
	PFS (U.S.)	6.00%	1.50%	7.41%
	PFS (non-U.S.)	7.00%	1.50%	8.40%

Figure 11: Losses in Production Process

4.3 *Identifying the Process Bottleneck*

In the spirit of the Theory of Constraints, it is worth pausing to perform step one in TOC, identifying the system's constraints (Goldratt 1999, 5). To simplify the process, let us ignore the market conditions and first only determine the internal constraint in the manufacturing process. Later chapters will then consider manufacturing and the market together to determine if the system is market constrained or capacity constrained.

The important attribute about Fluvirin is that it is manufactured on a seasonal schedule and changes throughout the year rather than operating continuously in a steady state fashion. Therefore it is important to look at the whole campaign instead of the available capacities at a single moment in time. Analyzing the entire season would take into account the capacities, lead-times, available inventories and operating windows.

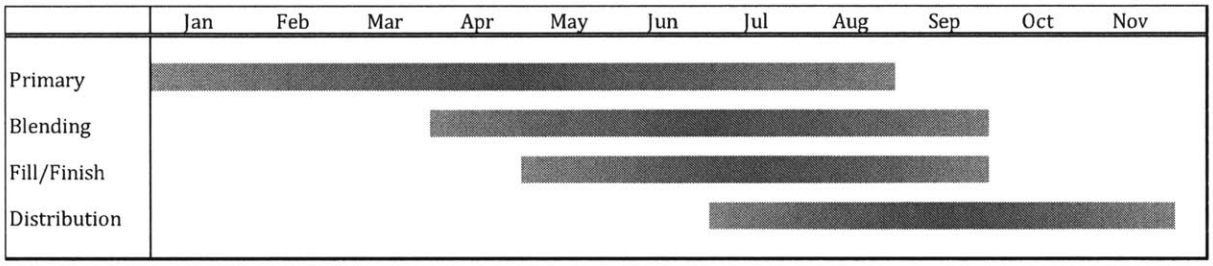

Figure 13: Campaign Production Timeline

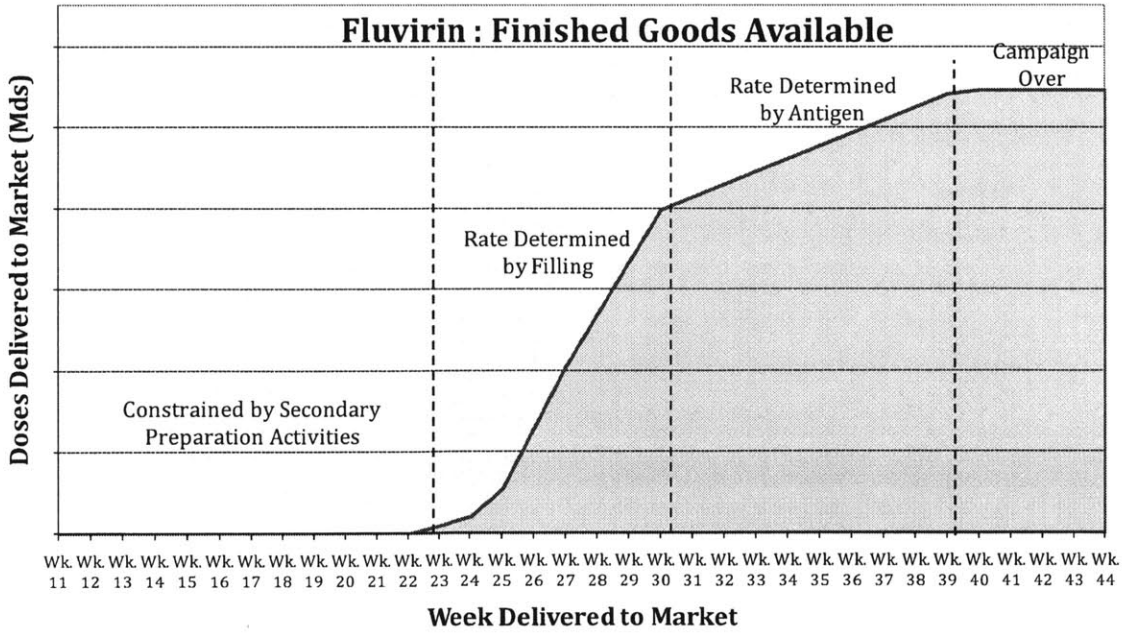

Figure 12: Finished Goods Delivered

As is seen in Figure 13 and Figure 12, a typical campaign has several stages. First, primary production begins building inventories of monovalents. Blending and filling then begin several weeks later and eventually begin delivering to the market in week 23. Between weeks 23 and 30, filling is operating at full capacity and is the rate that is limiting the final delivery to the market. By week 31 however, all of the monovalent inventories that can be blended have been consumed and the process is now limited by the amount of antigen that can be produced by primary. Blending and filling now only run occasionally when enough monovalent is available. This trend continues until the campaign window ends.

Primary production is therefore the process that limits the amount of vaccine that can be delivered to the market during the active selling season, thus making it the internal constraint. The specific point of the bottleneck in primary is the handoff of eggs between PIU and the inoculation step within upstream. PIU always has extra eggs available, but upstream is only able to accept a certain limit. If an egg is lost in PIU, primary will still be able to produce the same amount of HA because a replacement egg will be substituted. However, after the eggs are brought into upstream, the primary batch is established and no other eggs for that batch can be added. If an egg or fluid is then lost, primary will produce a batch with less HA than it would have if no losses had occurred.

Understanding the location of the internal bottleneck is critical for the subsequent focusing steps in the Theory of Constraints and correspondingly the results discussed in future chapters.

CHAPTER 5: MODEL DEVELOPMENT

Once the data for the main components is collected, each piece can now be connected together. Before this point in time, not only were there pieces of missing data, but the links did not exist. For example, it was unclear how valuable an increase in a strain yield (HA per egg of a particular strain) by 10% would be not only because the variable costs were unknown, but the relationship between strain yields, product delivery to patients and costs had not been defined. Taking a systems viewpoint allows us to understand those relationships across the entire value stream for both patient and financial purposes. We can start by thinking of how someone would perform the analysis for a single question guided by the concepts discussed in the *Haystack Syndrome* and then generalize it so that it is capable of answering our set of questions (Goldratt 1990, 72).

5.1 Manufacturing System Financial Structure

As can be seen in the Figure 14, the system definition incorporates all the key elements of the Fluvirin production environment (and in fact can be applied generally to many other settings). Focusing on a subset of this system could potentially lead to local optimization, so it is necessary to consider all of the components together. Although this diagram does not formally exist in the organization, it essentially represents the way in which the system interactions take place.

At the top of the diagram is the contribution margin, defined simply as revenues minus variable costs. (Throughput accounting similarly uses [Financial] "Throughput" as the key metric which is defined as revenues minus truly variable costs. In order to avoid confusion and to stay with the more conventional managerial accounting terminology, contribution margin will be used instead.) By maximizing the contribution margin of Fluvirin, we are also maximizing the profitability of the product. As has previously been discussed, the Fluvirin system is essentially independent of the rest of the Novartis network. Therefore, every increase that Fluvirin makes to its profitability leads to a 1:1 increase to the company. Operating expenses will be discussed at a later time.

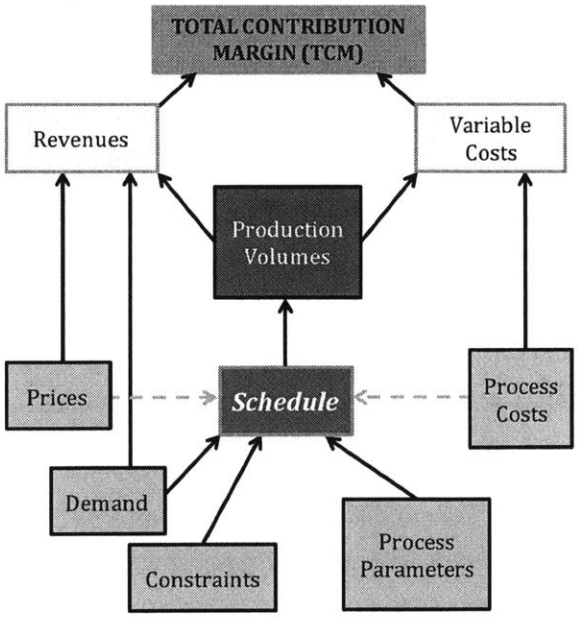

Figure 14: Simplified System Relationship

The left side of the diagram represents the revenues of the product by showing prices and quantities. Though the data collection chapters did not focus on the revenue side, it is necessary to have this information. In the case of the seasonal flu vaccine, most of the sales are derived from long term contracts with distributors, direct administrators and governments. The sales and marketing team provides the data for these contracted amounts as well as where there is uncertainty in their forecasts. Additionally, production volumes dictate revenues because a product must exist along with demand in order for a sale to occur.

On the right shows the variable costs as a function of the production volumes and process costs. The process costs are what were originally unclear in the organization, but are now defined because of the variable value stream cost mapping project in Chapter 3. These costs are defined by how they change in response to the production volumes at each process stage.

The center of the diagram shows the key link within the system: the schedule (Goldratt 1990, 117). The schedule dictates the production volumes and therefore the availability to the patients. It also drives the revenues and costs which then impact the contribution margin. Right now, the scheduling process is a relatively straightforward endeavour: make as much product as possible in order to try to meet demand. This objective is pursued by scheduling production levels at the point so that the material can flow as quickly as possible from start to end. When enough material is ready from one stage for the next to begin, it is started as soon as there is available capacity. Though there are challenges in the day to day activities of coordinating material so that the exact amounts are used, quality issues are closed and unexpected events are resolved, the high-level approach is pretty clear-cut. Unfortunately, due to the limited time window of the seasonal flu vaccine production system, it is not always possible to meet demand.

Below the schedule are the components that actually dictate when everything can occur. The constraints include typical scheduling limitations of capacity and timing. With Fluvirin, the production is not year-round, so some events can only start after certain dates. For example, the managers may decide that one strain cannot begin production until the World Health Organization (WHO) official announcement of which strains to use, but start prior to the announcement for one of the other strains. An additional constraint example is that since production is set at one batch a day, a maximum of 7 batches can be scheduled for a week. While the constraints place hard boundaries on the schedule, the process parameters establish how the schedule should be set. These parameters include strain yields, process losses and lead times. For example, since Fluvirin is a trivalent and therefore requires the antigen from all three strains, the schedule must be properly balanced. If one strain is much higher yielding than another strain, then it will be scheduled less frequently than the low yielding strain.

We can see from the diagram that changes at the bottom propagate all the way up to the contribution margin at the top. This begins to explain why it is so difficult to determine the financial impact of a change in the factory. To really know how a change impacts the company financially, it would require somebody to know all of these different pieces. Not only is it difficult to find this information because it is owned by several different groups, there is also a level of uncertainty with several parameters and some values were previously not known prior to the project (e.g. variable value stream costs). Once all the information is gathered and arranged, it then starts to become possible to perform a scenario analysis.

5.2 Scenario Testing

The first step in an analysis is to create a baseline assessment of the current system. Based on the constraints and process parameters, it is possible to determine how the schedule would be established in this situation. From the resulting production volumes and then financials, the contribution margin could be calculated for this baseline scenario. (Of course, the *exact* production volumes, revenues and costs cannot be determined, but the approximate values can be closely estimated from the data.)

If there is a proposal that we want to test, we would follow the exact same process steps to determine what the new scenario would look like. Comparing the baseline to the new scenario would then show the impact of the initial change in a "what-if" experiment (Goldratt 1990, 118).

Perhaps an engineer has validated an improvement that will reduce losses in primary production from 8% down to 6% and needs to determine if it is worth implementing the change to the factory. This proposal would be changing a process parameter. In this new scenario, that parameter change would necessitate a new schedule, the new schedule would create new production volumes, new revenues, new costs and hence a new contribution margin. (It is possible that sometimes the new values for these different components may be the same as the original values, but it is at least necessary to go through the steps.) The new production volumes can then be compared to the baseline production volumes in order to determine product availability to patients. Similarly, the new contribution margin can be compared to the baseline contribution margin in order to determine the financial impact.

If the rollout of the change required an investment, that can then be compared to the change in contribution margin to decide if the project should be pursued. As was mentioned before, operating expenses are not included in the model, but if a project is changing the operating expenses or creating a one-off expense or savings, that value can be compared to the impact on the rest of the system to inform the decision.

Despite the ease of talking through this example, there were several barriers that prevented this type of analysis from previously being performed. First, as was mentioned, there was an information availability challenge that has been overcome by the data collection and cost mapping. What currently would prevent this process from being run is the complexity in walking through these steps with the manual scheduling method. To really perform a reasonable analysis of a change, we would not only want to test a single scenario, but a range in order to capture the uncertainty in some of the values. Executing several sets of scenarios would simply be infeasible (and was inconceivable) in the current manual process because of the effort of making all of the scheduling decisions optimally while still meeting all the constraints and then calculating the resulting values. Rather than continue with the manual scheduling process, it would be preferable to automate those decisions.

As a substitute for the manual decision making process, we need a tool that will replicate the decisions as they would occur in these scenarios. Instead of looking at each decision individually and attempting to model those, we can replicate the goal of the planning team and the basic rules they follow in our tool. The rules are actually quite simple and have already been discussed: production can only be scheduled during valid times and material can only be consumed at a particular stage once it is ready. Also, the goal of the supply chain planning team is to make enough vaccine for patients and maximize the contribution margin. As a substitute for the human decisions, we can use an automated tool that also pursues an objective while following specific rules: a linear program.

Normally, linear programs are used when several combinations of tradeoffs can be made resulting in a huge number of possible options and its algorithm determines the best set of decisions to meet the objective, but we can use it in a slightly different way. For example, the linear program within a mapping website analyzes a combination of roads in order to make a recommendation to get a driver from point A to point B in the quickest time possible (Cormen 2001, 580). The algorithm for a linear program follows certain rules (only take valid roads) to pursue an objective (minimize time) based on certain parameters (distances and expected speeds). This is useful to someone who does not know and would have difficulty calculating the fastest route on their own. The piece of information that is valuable to the typical driver is *which route* to take. However, in this metaphor, we already currently know how to get from point A to point B. In this example, what we would want to know is the impact of a change. Consider the impact of adding a trailer to the car which would prevent it from travelling down certain roads. If the mapping program calculates again with a rule about the trailer, it can determine a new fastest route. By comparing the anticipated travel times between the original and new driving directions estimates, we can estimate the impact of the trailer on the travel times. For us, the valuable pieces of information are not the routes, but rather the *travel times* for each route.

Typically, the piece of information that is valuable to users of linear programs in a factory is what the production *schedule* should be (the route to maximum contribution margin) (Goldratt 1990, 118). With our model, the valuable piece of information provided by the linear program to us is actually not the schedule (because the planning team is ultimately going to work out those details) but rather the *contribution margins and production volumes* of the different decisions under consideration.

By walking through the process improvement proposal example again in more detail, we can see how the method works in Figure 15. The leftmost box includes a few high level process parameters that define the system (the real model contains several). To maximize the contribution margin, the linear program has determined the optimal schedule (second box). As can be seen in the third box, the schedule resulted in an outcome where not all the demand was able to be satisfied. (Only 50 million doses were manufactured despite demand of 60 million doses. Because it would have been preferable to fulfil all demand, there must have been a constraint on the system that prevented more vaccine from being produced.) Finally, the contribution margin for this baseline scenario is $50M based on the simulated revenues and costs (in Figure 16).

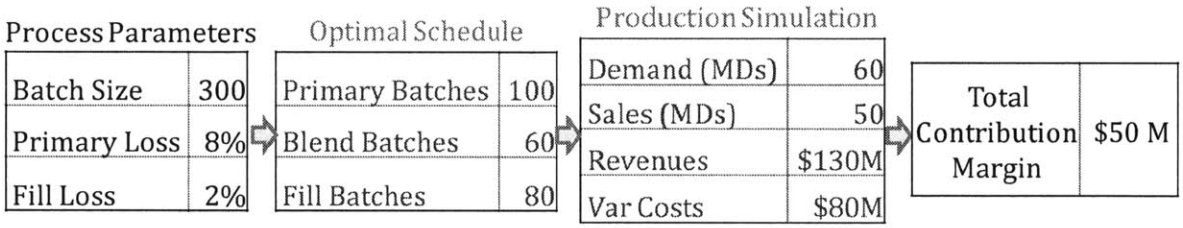

Figure 15: Baseline Scenario Example Calculations

	Unit Costs			Batch Costs			Total Costs
	Cost per Dose ($)	Number of Doses	Total Unit Costs ($)	Cost per Batch ($)	Number of Batches	Total Batch Costs ($)	Total Costs ($)
Primary	0.06	50,000,000	3,000,000	94,500	100	9,450,000	12,450,000
Blending	0.05	50,000,000	2,500,000	10,000	60	600,000	3,100,000
Fill/Finish	1.23	50,000,000	61,250,000	40,000	80	3,200,000	64,450,000
Total			66,750,000			13,250,000	**80,000,000**

Figure 16: Baseline Scenario Costs

When considering the proposal, the model is adjusted so that the primary losses have now been reduced to 6% in Figure 17. A fresh execution of the linear program shows the outcome of the change. What the process improvement accomplishes is that 2 million more doses could be produced in the same number of batches in primary. As a result, additional effort needed to be scheduled for the blending and filling stages. (Since it is clear that production levels in blending and filling can be increased, the constraint in the original scenario was the 100 batches that could be scheduled in primary.) The new increased production quantities therefore led to an increase in revenues. However, because additional effort would be needed in some areas, costs also rose. Our new contribution margin is $52M up by $2M over the baseline. We could then compare the $2M benefit to the project costs directly for short-term decisions or through NPV analysis for long-term decisions.

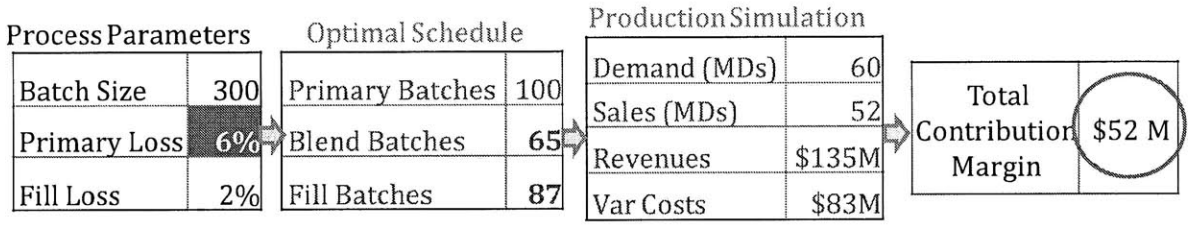

Figure 17: Scenario Updated with Primary Improvement

	Unit Costs			Batch Costs			Total Costs
	Cost per Dose ($)	Number of Doses	Total Unit Costs ($)	Cost per Batch ($)	Number of Batches	Total Batch Costs ($)	Total Costs ($)
Primary	0.06	52,000,000	3,120,000	94,500	100	9,450,000	12,570,000
Blending	0.05	52,000,000	2,600,000	10,000	65	650,000	3,250,000
Fill/Finish	1.23	52,000,000	63,700,000	40,000	87	3,480,000	67,180,000
Total			69,420,000			13,580,000	**83,000,000**

Figure 18: Scenario Costs with Primary Improvement

This example again highlights why it was essential to split the process costs into the separate components. Between the two scenarios, only some costs actually changed in order to produce the additional doses. Due to the cost mapping, we know there will be an increase in unit costs to make the additional 2M doses as well as an increase in batch costs for running additional batches in blend and fill as can be seen in Figure 16 and Figure 18. Within the primary production step however, the batch variable costs stayed exactly the same between both scenarios. If the primary variable batch costs compose the heavy majority of costs throughout the value stream, there would have been very little cost impact due to the proposed change. Conversely, if the majority of variable costs were in blending or filling, then there would have been a large rise in the total variable costs to produce the additional doses. Simply using a single average dose cost would have distorted the cost impact of this proposal.

The example also illustrates a key aspect of the model's behavior. In this situation, due to the unmet demand, it was preferable to use the improved production methodology to increase production levels. However, if the system was not capacity constrained, but rather market constrained and there was no reasonable expectation that there would be any additional demand, the model would have behaved differently. Instead, the simulation would have used the increased efficiency in primary to make the necessary amount of doses to meet demand. Higher efficiency in primary would require fewer batches to be produced in that location. To still meet demand, the same number of batches would still need to be scheduled through blend and fill, so those costs would stay the same. Under this new scenario, we again see that the cost structure would not stay at a fixed ratio. A value stream cost map shows us that the primary costs would be reduced while blending and filling costs would remain constant. Relying on a single cost for the doses would have again distorted reality from the true economic changes occurring.

5.3 Linear Program Formulas

To understand how the automated decision making would be structured, we must understand the three components of a linear program: the objective function, decision variables and constraints. The following notation will be used:

i Week number ϵ {1, 2, ... 52}

j Process step ϵ {primary, blend, fill, sale}

k Material type ϵ { {primary type}, {blend type}, {fill type}, {final product sale type}}

Primary Type ϵ {H1N1, H3N2, B}

Blend Type ϵ {Trivalent for Vials, Trivalent for U.S. PFS, Trivalent for non-U.S. PFS}

Fill Type ϵ {Vial, U.S. PFS, non-U.S. PFS}

Final Product Sale Type ϵ {Vial, U.S. PFS, non-U.S. PFS}

X_{ijk} Number of batches scheduled in week i for process step j for material type k

R_{ik} Revenue in week i of material type k (final product sale type)

C_{ijk} Cost in week i for process step j for material type k

Y_{ijk} Cumulative remaining inventory for week i in process step j of material k

Z_{ijk} Amount of material started in week i for process step j using material k

B_{ijk} Batch size (number of doses) in week i for process step j for material type k

l_{ijk} Material loss (process and failure) in week i for process step j for material type k

t_{jk} Lead-time for process step j for material type k

N_{ik} New demand for week i of material k (final product sale type)

D_{ik} Cumulative remaining demand for week i of materials k (final product sale type)

$c_{unit,ijk}$ Unit cost in week i for process step j for material type k

$c_{batch,ijk}$ Batch cost in week i for process step j for material type k

p_{ik} Expected price in week i of material type k (final product sale type)

5.3.1 Decision Variables

Within the linear program exists the variables that the algorithm is deciding on in order to maximize the objective function. In the model, the values relating to the schedule are the decision variables. The schedule decision variables answer the question: how much, of what, when? If the program decides to schedule 5 batches of strain B during week 12 for primary, it would set the variable to $X_{12,\text{primary,B}}=5$.

$$Decision\ Variable: \quad X_{ijk}$$

To capture the entire system behavior, it is important not only to have a production schedule but also a sales schedule which would work in the same way. Revenue will be generated when sales are scheduled (e.g. 100,000 vials sold in week 30: $X_{30,\text{ sale, vial}} = 100{,}000$).

5.3.2 Objective Function

The objective function is the component of a linear program that the algorithm pursues (it can maximize or minimize). The Fluvirin model will focus on maximizing the contribution margin and consequently the production volumes will be maximized to meet patient demand. As stated earlier, the contribution margin is simply defined as revenues minus variable costs.

$$Maximize\ Contribution\ Margin: \quad \sum_{ik} R_{ik} - \sum_{ijk} C_{ijk}$$

5.3.3 Constraints

Each rule that the linear program must follow is defined as a constraint. The first rule about scheduling during valid times places a constraint on the maximum number of batches that can be scheduled during a period of time. If primary production can run seven batches in one week, then the sum of all of the batches that are scheduled in primary must not exceed seven. Additionally, if production is not allowed for a particular strain or process during part of the year, the maximum that can be scheduled during that period would simply be 0.

$$\sum_{k} X_{ijk} \leq Maximum\ Process\ Step\ Schedule_{i,j} \quad \forall i,j$$

$$X_{ijk} \leq Maximum\ Material\ Type\ Schedule_{i,j,k} \quad \forall i,j,k$$

The second rule is that material cannot be consumed by one process until a sufficient amount has been produced by the previous process. This can be implemented mathematically in the model by requiring that the inventory quantity between two stages must always be greater than or equal to 0. Sales are regulated through the same rule as well. For a sale to be scheduled, enough finished goods must exist in order for units to be consumed in the subsequent process step (i.e. being sold in the market). Additionally, not only are finished goods materials consumed when a sale occurs, but also demand is consumed by the process (i.e. each sale reduces the inventory of finished goods by one as well as the "inventory" of unmet demand). Just as with inventories of materials, the inventory of unmet demand must also be greater than or equal to 0 (otherwise the program would continue to sell when there were no customers).

$$Y_{ijk} \geq 0 \quad \forall i,j,k$$

$$D_{ik} \geq 0 \quad \forall i,k \in \{\{\text{final product sale type}\}\}$$

5.3.4 Supporting Functions

Combining the objective function, decision variables and constraints defines the linear program structure for Fluvirin. The rest of the links in the diagram connect up to form the rest of the model and must also be defined. When the process parameter batch data (yields, losses, etc.) are combined with the scheduling variables the forecasted production volumes are created. Merging the schedule and production

volumes with the value stream costs creates the variable costs. Finally, multiplying the sales schedule with the prices creates the revenues. When executed, the linear program simply determines the optimal schedule in order to maximize the contribution margin.

Defining the cumulative inventory is rather straightforward. It equals the inventory from the previous period minus the material that is being consumed by the next process plus the material that was just completed for this process (delayed by lead-time t of when it started in the previous step). The material completed during the existing period is also scaled to account for any losses in the process.

$$Y_{i=1,j,k} = 0 \quad \forall j,k\text{: } Inventory\ starts\ at\ 0\ for\ week\ 1$$

$$Y_{ijk} = Y_{i-1,j,k} - Z_{i,j+1,k} + (Z_{i-t,j,k} * (1 - l_{ijk})) \quad \forall i,j,k$$

Similarly, the cumulative demand can be represented in a parallel fashion. It equals the remaining demand from the previous period plus the new demand minus the demand that is now being satisfied by means of a sale.

$$D_{i=1,k} = 0 \quad \forall\, k \in \{\{\text{final product sale type}\}\}\text{: } Cumulative\ demand\ starts\ at\ 0\ for\ week\ 1$$

$$D_{i,k} = D_{i-1,k} + N_{i,k} - Z_{i,sale,k} \quad \forall i,k \in \{\{\text{final product sale type}\}\}$$

The amount of material started in each process is a function of the schedule and batch sizes.

$$Z_{ijk} = X_{ijk} * B_{ijk} \quad \forall i,j,k$$

Total costs are comprised of the batch and unit costs for each step and material.

$$C_{ijk} = (c_{batch,ijk} * X_{ijk}) + (c_{unit,ijk} * Z_{ijk}) \quad \forall i,j,k$$

Revenues are generated for each sale that occurs.

$$R_{ik} = p_{ik} * X_{i,sale,k} \quad \forall i,k \in \{\{\text{final product sale type}\}\}$$

Some batch sizes are fixed quantities based on the operating specifications. For example, each syringe filling batch runs for 250K doses ($B_{i,fill,k} = 250{,}000$). Batch sizes in primary production (j=primary) however, are based on several process parameters:

$Egg\ Quantity_i$	Number of eggs started in the harvesting process (# eggs)
$Harvest\ Yield_i$	% of good eggs started in primary production (%)
$HA\ per\ Egg_{ik}$	HA captured from each harvested egg (ug of HA/egg) (commonly called strain yield)
$Blend\ Target_{ik}$	Amount of monovalent blended into trivalent (ug HA/mL)
$Fill\ Quantity_i$	Dosage of trivalent (mL/dose)
$Overfill_i$	Amount of extra trivalent added during filling to ensure minimum requirements (%)

Each variable is also specific to the week i, though most will remain constant throughout the campaign. The amount of HA per egg and blend targets are specific to the material type k, since this is for j=primary, the material types are the different strains: H1N1, H3N2, B. Because these values are different for each strain, each strain yields a different number of doses coming out of primary.

$$B_{i,primary,k} = \frac{Egg\ Quantity_i * Harvest\ Yield_i * HA\ per\ Egg_{ik}}{Blend\ Target_{ik} * Fill\ Quantity_i * Overfill_i}$$

5.4 Summary

Figure 19 has updated the original generalized figure to be specific to Fluvirin. Each solid box corresponds to a separate part of the model (created in Excel using the upgraded Risk Solver Platform) that represents the system. Based on the process parameters and constraints for each part of the manufacturing system, production is scheduled. The production simulation creates "doses" (Z_{ijk}) that flow from primary to blending to filling and then into the market. Each execution of the simulation shows the new production volumes and contribution margins.

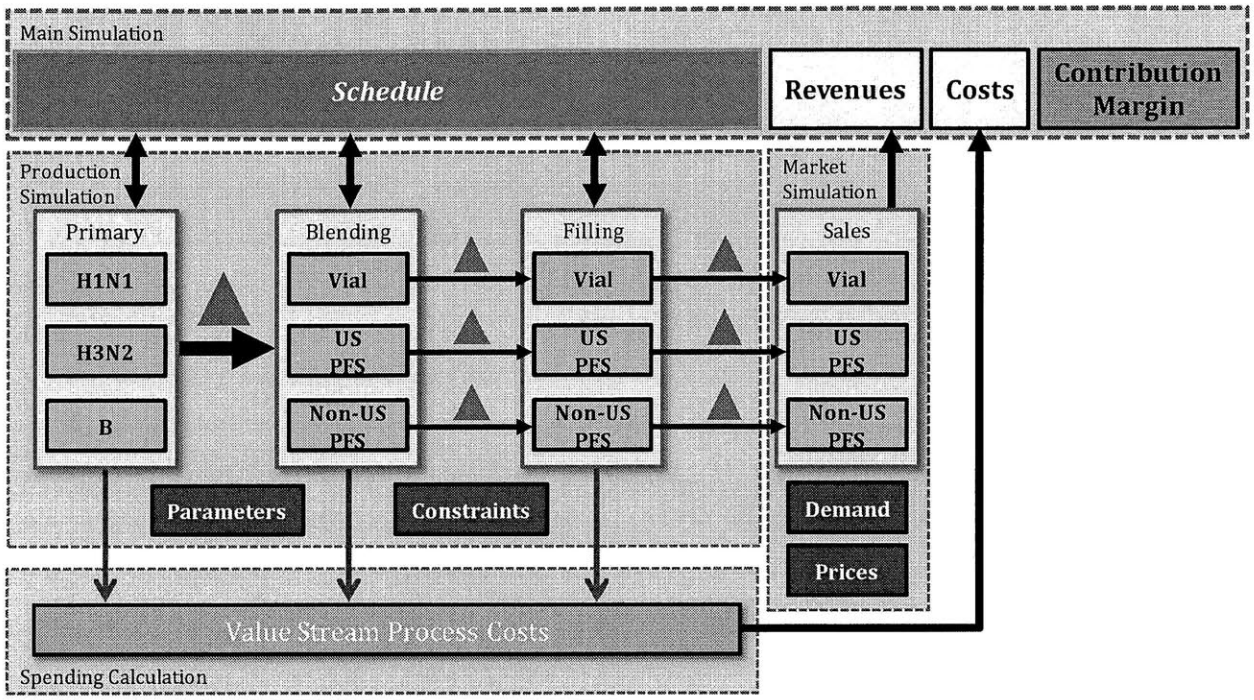

Figure 19: Fluvirin Model Architecture

Once the model is constructed, we can begin to see how different potential changes would affect delivery to patients and the contribution margin. Figure 20 lists 18 different types of changes that can be tested in the model in order to understand the impact. Some changes, such as negotiating lower prices with suppliers are already easy to calculate; but other changes, such as increasing the seed yield for a strain or improving the processing methods are indirectly tied to the results that their impact had been unknown. These 18 categories of changes are at the high-level, each can drive different parameters and in through several different ways. For example, increasing strain yield can be achieved by increasing the strain yield for strain B or for H1N1 and can be achieved by spending more time in the lab or switching to a new seed. It would be great if all of these changes could be implemented simultaneously and without any cost, but financial and personnel resources are limited. Using the model will now allow us to identify which projects are cost effective and how they should be prioritized.

Improvement Being Considered	Example of Implementation	Direct Result on Parameter	Result for Customer	Impact on Contribution Margin
HA per Egg	Increasing seed yield	Increase primary batch size: $B_{i,primary,k}$	Increased product availability, earlier delivery	Increase in sales (given sufficient demand), Lower average costs due to increased efficiency
	Improving processing methods			
	Purchasing larger eggs			
Blend Targets	Increasing strain stability			
Egg Quantities	Expanding egg processing capacity			
Harvest Yield	Improving screening process during candling			
	Receiving better eggs from farm			
Process Losses	Reducing physical losses due to errors through better practices	Reduction in process losses for each step: l_{ijk}		
	Shortening processing time to reduce risk of exceeding discard time			
Blending or Filling Batch Size	Adding capacity to the lines	Increase secondary batch sizes: $B_{i,blend,k}$ or $B_{i,fill,k}$		
	Lengthening run-times			
Lead-times	Reducing supply-chain timelines	Shorter time between start and finish of batch: t_{ijk}	Earlier delivery	Earlier production helps guarantee sales raising expected contribution
	Shortening testing lengths with better technology			
Production Availability	Shortening cycle times to be able to run more frequently (primary)	Increase in the constraints: Maximum Process Step Schedule or Maximum Material Type Schedule		
	Running during previously idle periods (secondary)			
	Starting manufacturing before official announcements			
Value Stream Costs	Reducing wasted materials	Reduction in costs: $c_{unit,ijk}$ or $c_{batch,ijk}$	No direct impact	Lower average costs
	Negotiating lower prices with suppliers			

Figure 20: Relationship Between Project, Model Parameters and Impact

Finally, the model is not simply intended for the operations teams. With an improved understanding of the marginal costs and operational capabilities, the sales team can be better equipped when making decisions regarding new contracts. The team will have a better idea of the potential profitability of entering new markets and be able to test out when new products could be delivered to patients.

CHAPTER 6: MODEL ANALYSIS AND APPLICATION

This chapter will discuss the validation of the model and the various ways it can be applied to help with decision analysis. The initial concept is based off of a single cost-benefit analysis, but many more techniques can be applied. Making decisions in the face of uncertainty, understanding the sensitivity of each parameter on the contribution margin and prioritizing projects can all be derived from the existing the model. Using these new tools will improve the quantitative aspects of these decisions from a strategic down to an operational level and will also improve intuition for those decisions.

6.1 Validating and Updating the Model

Before immediately applying the model, it is necessary to validate that it performs as expected. The key assumption in using the linear program is that it will be replicating the decisions that the planning team would make if they faced the same constraints and parameters. Executing the model for previous years (using the parameters and constraints for those years) and comparing it to the actual schedule and production volumes shows how representative the model is of the real system.

Upon performing the comparison, one main difference can be seen that needs to be addressed. While it was originally said that we are running the linear program in order to calculate the contribution margin and the actual schedule details were ancillary, it would be nice to get a result that had the maximum contribution margin with the exact schedule that would be expected. For example, in a year with an extra week of capacity, the real planning team would start at the beginning of the year and not plan on using the final week of the campaign for production (this gets the product out the door earlier and provides flexibility in the case of errors or increased demand). Since the model was only trying to maximize the contribution margin, the linear program was originally indifferent to what the schedule actually was. In this case, it would arbitrarily have put the open week during any part of the year. In order to encourage the program to perform consistently with the planners, prices that were originally the same between consecutive weeks can be slightly biased so that earlier delivery is promoted. Setting a price \$.0001 lower for each subsequent week will have a negligible impact on the contribution margin, but will cause the schedule to deliver as early as possible and thus more in line with the actually plan.

After the model is verified with historical plans, the automated schedule for the upcoming campaign can be approved by the planning team. Some values will inevitably need to change over time as managerial decisions evolve or process estimates about the upcoming year are revised as more data is generated. Constraints in the model corresponding to when certain strains can begin production and the maximum number of batches that can be run can easily be modified by the planning team if those details

change over time. Process parameters about yields and such can be updated by the operations teams, demand values by the sales teams and cost data by the finance team.

6.2 Example Execution

A simple execution of the model using fictional data was run to show some of the outputs. The appendix contains several screenshots of the various components of the Excel model. Figure 21 shows the expected output from primary production. It tracks the number of doses that would be available for blending of each monovalent. Each strain is started during a different week because of the constraints on the schedule. Also, because each strain has a different yield, inventories of the higher yielding strains build faster than the lower yielding strains. Finally, near the end of the campaign, the schedule is balanced so that the same numbers of doses are always available for blending to begin.

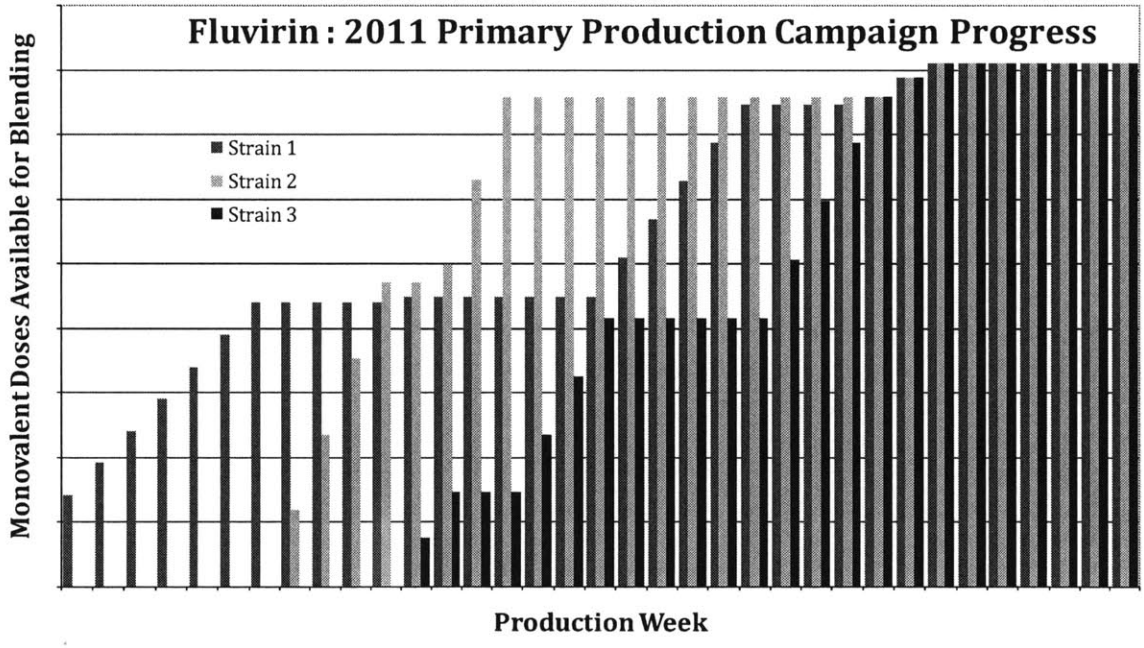

Figure 21: Fluvirin Baseline Scenario Cumulative Inventory of Monovalents (Hypothetical)

Figure 22 shows the number of final doses that are projected to be ready during each week. Based on the existing contracts and forecasted demand, different quantities of each type of presentation are planned.

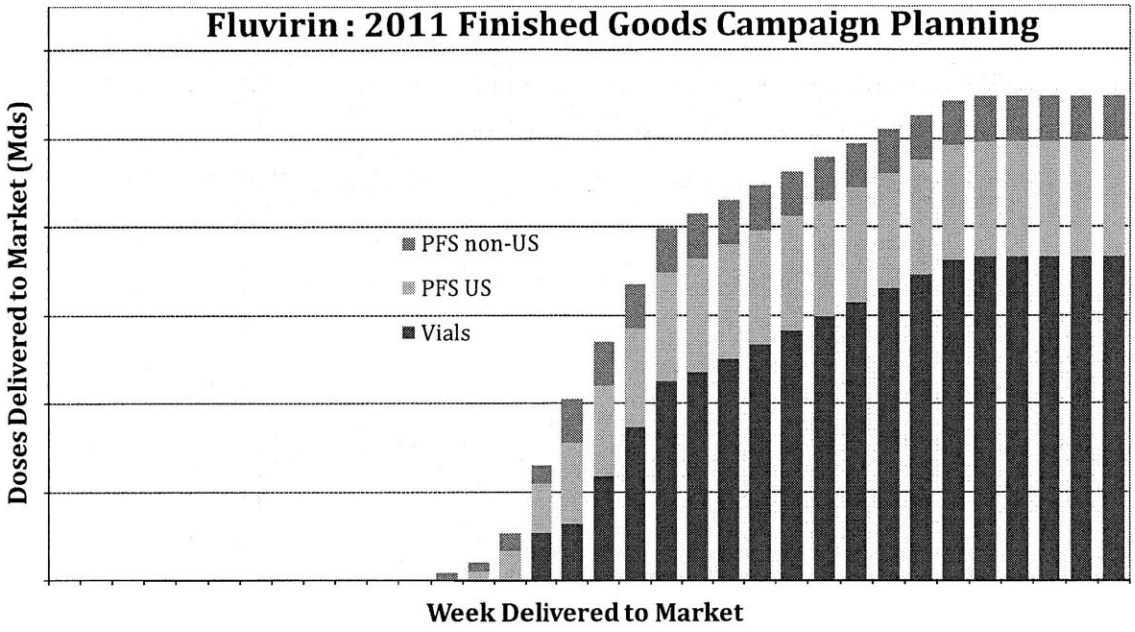

Figure 22: Fluvirin Baseline Scenario Finished Goods Delivered to Market (Hypothetical)

In order to determine the cost impact of the scenario, the value stream costs are multiplied by the scheduled values to determine the total costs (Figure 23). The model only tracks aggregate costs for each stage, but managers can easily dig into the underlying data to better understand the values and make updates when needed.

	Unit Costs			Batch Costs			Total Costs
	Cost per Dose ($)	Number of Doses	Total Unit Costs ($)	Cost per Batch ($)	Number of Batches	Total Batch Costs ($)	Total Costs ($)
Primary	0.25	60,000,000	15,000,000	850,000	200	170,000,000	185,000,000
Blending	0.20	60,000,000	12,000,000	150,000	60	9,000,000	21,000,000
Fill/Finish	2.80	60,000,000	168,000,000	120,000	60	7,200,000	175,200,000
Total			195,000,000			186,200,000	**381,200,000**

Figure 23: Fluvirin Baseline Scenario Cost Structure (Hypothetical)

6.3 Scenario Testing Including Uncertainty

The original examples for running the model were for analyzing a cost-benefit decision where all variables were assumed to be known. To develop a more robust approach, we must consider the decision while recognizing the inherent uncertainty with forecasting the system variables.

First, we should acknowledge that there are two different types of uncertainty in the model based on the timing of the unknown value. For example, it is not currently known exactly what the yield will be for the flu strains for a future year. Though the strain yield is not currently known, it will be known during the time of the actual scheduling decisions. The uncertainty will therefore be resolved prior to finalizing the scheduling decision. If one strain yield turned out to be higher than originally expected and one yield was lower, the schedule would just be adjusted to rebalance levels. By reacting, the schedule always represents the best that can be achieved for that situation. Conversely, some pieces of the system may require the schedule to be locked in before the uncertainty is resolved. If the schedule had to be frozen before the yields were determined, the outcome would likely not be optimal for those conditions (i.e., too many doses of one strain would be produced and too few of another strain would be produced). Demand fits into both categories of uncertainty. When the early planning is being done, not all contract agreements have been finalized. The contracted demand will be resolved by the time scheduling is taking place. Additional quantities ordered or returned products however, will not be known until after the production has occurred. With the first type of uncertainty, the parameters should be changed and then generate the optimal schedule. In the second type of uncertainty, the schedule would be generated based on the initial data, but then the final results would be based on the updated data.

In order to test the model for uncertainty, a range of values can be used to witness how the system would react. Since the exact baseline situation cannot be known with absolute certainty, we can analyze a range of values around which it is believed the likely conditions will exist. Along with performing the scenario analysis for the original example of comparing the improvement from 8% to 6% in the standard condition, other situations which could potentially occur should be considered. Based on historical data analysis, it is possible to identify which parameters tend to be more uncertain than others and determine a range of values which they may take on, thus creating a region of likely values. Since there is uncertainty about strain yields, the same comparison can be run when yields are lower and the change is implemented. Or, because demand is unpredictable, comparison can be run with a lower demand value for the 8% and 6% scenarios. When performing this analysis, it is critical to always compare like for like. Comparing a high demand scenario with the 8% process losses to a low yield scenario with the 6% process losses would not be valid because that would not represent two parallel scenarios.

Many of the decisions are simply accept or reject. (For the sake of simplicity in this example, assume that a project will be accepted if the new contribution margin is greater than the baseline and reject otherwise.) If the result of each parallel scenario comparison is the same for the range of parameters being considered, then the decision can be implemented with a high level of certainty. As illustrated in Figure 25, every decision at each corner is to accept the proposed change. Otherwise, if the result is unfortunately inconsistent depending on which set of uncertain values are being used, as illustrated in Figure 24, further effort would need to be taken to improve the forecast (which would tighten the range of values) or mitigate the risk of the parameter (which would lessen the severity of the impact if the invalid decision was pursued).

Corners		Contribution Margin		
Demand	Yield	Baseline	Proposal (Includes $1M transition cost)	Decision
Low	Low	$49M	$52M - $1M	Accept
Low	High	$51M	$54M - $1M	Accept
High	Low	$49M	$54M - $1M	Accept
High	High	$52M	$57M - $1M	Accept

Figure 25: Consistent Corner Case Result

Corners		Contribution Margin		
Demand	Yield	Baseline	Proposal (Includes $1M transition cost)	Decision
Low	Low	$48M	$51M - $1M	Accept
Low	High	$52M	$52M - $1M	Reject
High	Low	$48M	$53M - $1M	Accept
High	High	$53M	$55M - $1M	Accept

Figure 24: Inconsistent Corner Case Result

6.4 *Sensitivity Analysis and Parameter Sweeps*

Each key parameter to the model can be tested across its range of possible values to observe how significantly it affects the contribution margin. Knowing the sensitivity will show how much a parameter would need to move in order to achieve a desired result and develop intuition as to why the system would react in a certain way based on a change. Sensitivity analysis data is collected by holding the other parameter values constant and sweeping the variable from the low to the high values. An optimization is executed and the resulting contribution margin is saved for each sample value. The larger the change in the contribution margin for each increase of the parameter, the greater leverage that parameter has on the overall system. Knowing the leverage for each variable is crucial for two reasons. First, for variables that are uncertain, it helps to establish which variables could potentially distort the model and therefore lead to an incorrect decision. Those parameters could then be further analyzed to increase the accuracy of the value and thus improve the fidelity of the model. Secondly, identifying the high leverage variables can help managers identify where to focus their resources on the areas that would provide the greatest returns.

An important aspect of the sensitivity analysis is that it can very often be non-linear. For example, let's analyze the impact of a parameter on the contribution margin in Figure 26. In this case, the parameter relates to the number of doses being produced in each primary batch. (Many of the parameters in the Fluvirin model directly influence the batch yield including process losses, strain yields, egg quantities, blend targets, etc.) Under the batch a day approach, a larger batch size also means increased throughput for that stage. As can be seen in the figure, when the number of doses produced in each batch (i.e. the throughput of the primary stage) increases, the contribution margin increases. However, there is a bend in the curve where the parameter changes from being a high leverage variable to having less leverage on the contribution margin.

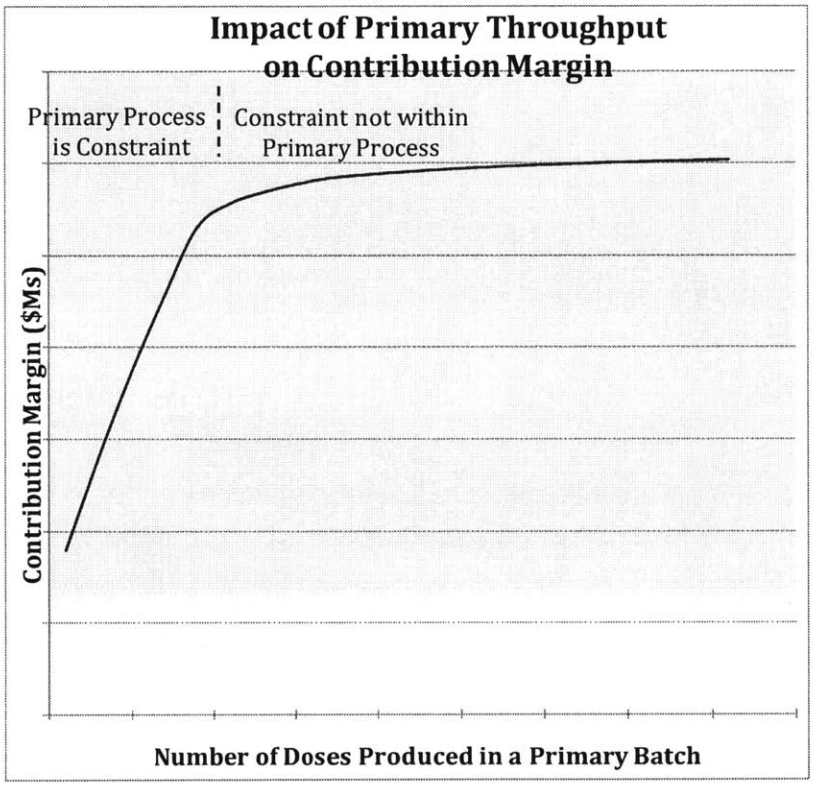

Figure 26: Response of Contribution Margin by Primary

The underlying reason for this transition is explained by the Theory of Constraints. When the number of doses produced per batch is low (in the high leverage region) the primary manufacturing process throughput is so low that it is the system constraint (i.e. the bottleneck). Every increase in the primary batch size increases the throughput of the system and allows for additional doses to be sold, thus the significant rise in the contribution margin. Eventually, as the number of doses produced increases,

primary suddenly is no longer the constraint and a different part of the system is constraining its performance. The other constraint in the system could be internal, such as the blend or fill capacities, or it could external, such as the market demand. Without this other subsequent constraint, the increases in the contribution margin would have continued on their original trajectory. Now, increased throughput in primary would not translate into an additional sale because of the constraint. Instead, once the primary process constraint is broken and the transition occurs, the additional minor increases in the contribution margin are due to the increased efficiency in the batch production. Just as the supply chain team would do, the model opts to make the maximum number of doses needed, but now with the increased throughput, using fewer batches. Since there are batch costs in the Fluvirin process, reducing the number of batches running reduces the variable costs and thus increases the contribution margin.

In systems where there are no batch costs, the curve would look different. This is what is typical in many manufacturing settings where there are predominantly unit costs. For those environments, increasing the batch size or throughput at a non-bottleneck has no financial benefit because not only does it not lead to more system throughput, it consumes the same amount of variable costs. Car production for example, which has high unit costs and very few batch costs, would not benefit from increased throughput at the non-bottleneck because it would still consume the same amount of materials (an extra set of doors and tires, an additional engine, etc.). Systems with low or no batch costs would have a very flat curve in the region after the constraint is broken for that process. Due to the considerable transition between the constrained and un-constrained regions, management must therefore know which region each process is in so that the proper levels of investment can take place.

Conversely, if variable batch costs are extremely high relative to the unit costs and prices, the slope in the region of the graph after the constraint is broken remains high (although lower than the region in which the process is the constraint). For high batch cost environments that have low margins, increasing throughput by increasing the batch size saves almost as much as selling additional units. Knowing the exact status of that process then becomes less important because it would always be worthwhile to continue to increase batch size whether or not it is the constraint. In this type of production environment, the conclusion deviates from traditional Theory of Constraints thinking because it would still be worthwhile increasing throughput at a location other than the constraint.

Using the model not only helps quantify the financial impact of changes, but also helps identify where the constraints in the system would exist. Managers can then know which parameters to focus on, how much to invest and how much of an improvement is needed to have the most significant impact.

6.5 Parameter Evaluation for Project Prioritization

The initial baseline scenario testing and corner case testing were designed to assist decisions in significant changes to the system like buying a new piece of equipment. In that analysis, it is critical to know the absolute magnitude of the change to determine whether an investment is worthwhile. For smaller projects however, which often require minimal spending, knowing the relative magnitude is more important for knowing where to focus efforts. Due to the uncertainty associated with the different cost elements and the indirect non-linear connections between different parameters and the contribution margin, the site is looking for an improved way of knowing what to work on to achieve the greatest results.

Evaluating process improvement projects is typically performed by coming up with a few ideas, assessing each one's value and then prioritizing them. In addition to the already mentioned challenge of assigning a value, this approach also leaves open the possibility that some projects are never thought of because of existing mental models about where to focus efforts. Basically, if everyone thinks the important area to focus on is in a particular location, that is where the majority of ideas will address. This potentially misses out on projects that might exist in the areas that are undervalued by the group's mental model.

By using the model, not only can we more accurately evaluate decisions, but now we are able to change the way in which proposals are solicited. Rather than think of projects and then evaluate how they impact the contribution margin, we can use the model to assess what impacts the contribution margin the most and then think of how to implement the projects that create that impact. This approach reverses the methodology in which ideas were pushed to now being pulled based on the data.

To identify which types of projects most significantly impact the contribution margin, we need to identify which parameters affect the contribution margin. We can change each parameter by a small amount one at a time from the baseline to see how the much contribution margin changes. For example, by testing the impact of improving each parameter by 1% we will see which are most important. Though it would be inappropriate to linearly extrapolate the results of this (e.g. by multiplying the result of the 1% change by 20 in order to estimate the impact of a 20% change) it is reasonable to use this value as a guide for the small changes that would result from process improvement projects. Once all of the parameter data is collected, managers can then identify which variables affect the contribution margin the most. They can then petition for ideas within their groups to come up with improvements that target those parameters. Finally, they can weigh the financial impact from the model, complexity of each project and the potential magnitude of the parameter change to prioritize the new projects.

CHAPTER 7: CONCLUSIONS AND RECOMMENDATIONS

7.1 Summary of Results

7.1.1 Capital Investments

Due to the confidential nature of the work, the specific projects for capital investments will not be discussed. However, the previously discussed techniques of using corner cases to understand the overall financial impact of a decision in the face of uncertainty were used.

7.1.2 Production and Sales Planning

Managers from the operations and sales teams are now able to quickly determine the impact of different production and sales decisions. This helps when trying to make decisions regarding the prices and quantities to set in contracts.

With more accurate cost information that captures the real cost of producing additional doses, rather than the overestimates based on standard costs or underestimates based on direct materials costs, managers can determine how to handle instances of demand uncertainty using the newsvendor model by finding the optimal probabilistic tradeoff between overage costs and opportunity costs (Beckman and Rosenfield 2008, 147).

Production scenarios, such as the example in Figure 27, were also played out to demonstrate when the vaccines would be available. This allows the operations team to illustrate to the sales managers how various changes would affect when doses would become available. Data is published by the Centers for Disease Control on the quantity of doses distributed in each week of a campaign so that managers can determine if their plans will help deliver to the market earlier than the competition historically has. When analyzing the graphs, instances of when contracts cannot be fulfilled are easily identified and corrective action can take place. Conversely, during times when excess capacity exists, additional contracts can be evaluated to help generate additional sales.

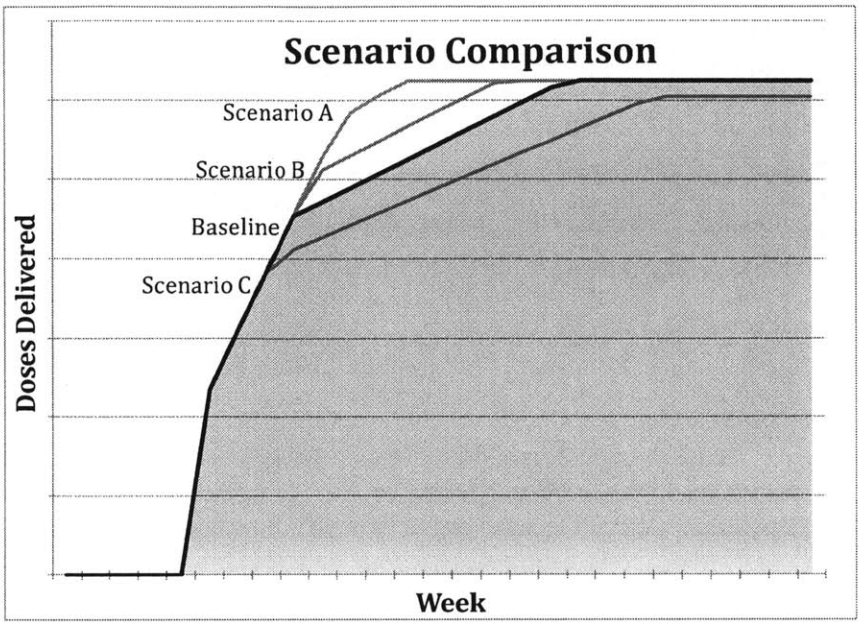

Figure 27: Delivery Schedule of Various Scenarios

Finally, the sales team also needs to understand the potential profitability when considering new markets because of the increased costs associated with the additional venture. Regulatory approvals and product modifications (e.g. designing a label in a new language) require a substantial amount of effort and cost. Though they were previously reluctant to make many expenditures without knowing the real costs, having an increased confidence in the accuracy of the cost data will now allow for more opportunities to be considered.

7.1.3 Process Improvement Project Evaluation

In order to help focus process improvement projects in the most effective areas, the different parameters within the model were tested around the baseline condition in order to calculate their leverage. Each parameter was tested as a 1% improvement from the existing state, though the actual expected improvement would need to be scaled appropriately to make direct comparisons between projects. To demonstrate the important difference between when the production system is market constrained and capacity constrained, two different executions of the model were created.

First, when the site is already able to produce a sufficient number of doses (i.e. market constrained), the clear priority is to drive sales. The remaining parameter changes are all efficiency improvements that reduce the total costs of meeting the demand. The importance of increasing sales in and of itself is not a dramatic insight since it is the only approach that actually increases the constraint.

However, the relative scale seemed to surprise managers. Because the cost structure data had not previously existed, the only number that was ever used to calculate potential changes in sales was the standard cost. Since standard costs include allocated fixed costs, the potential increase in profitability was underestimated. As a result of this underestimate, too much emphasis focused inward on reducing operations costs instead of focusing on the market. Once the data was collected in Chapter 3 to distinguish the different costs, the real value could be calculated.

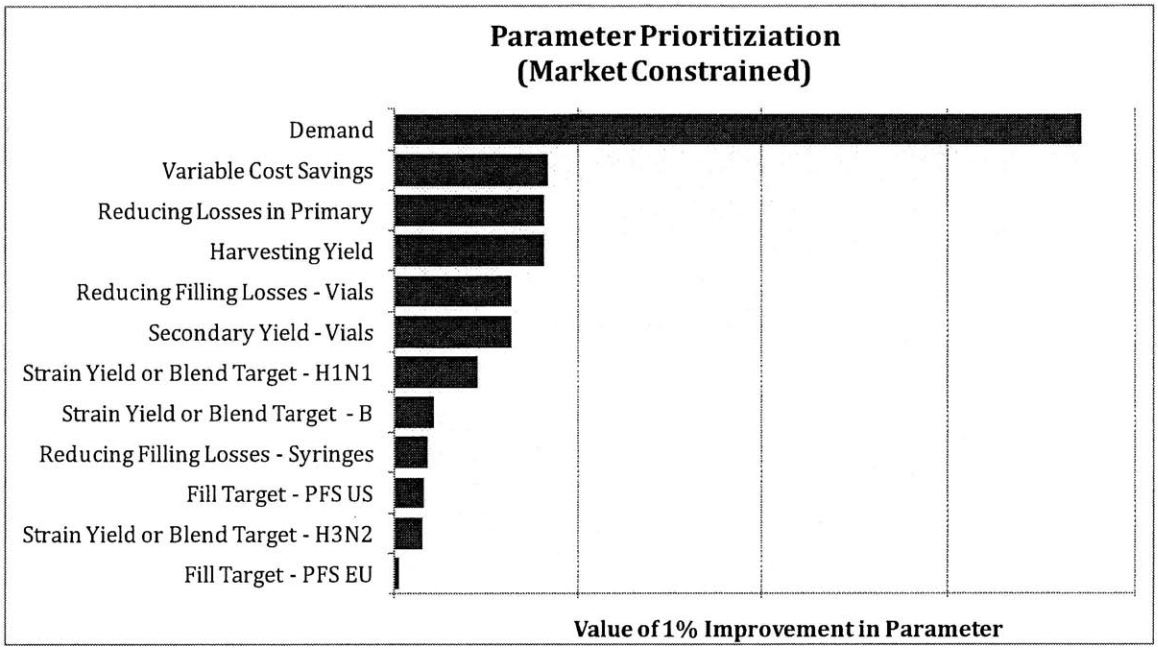

Figure 28: Parameter Prioritization During Market Constrained Season

This massive difference between the importance of growing sales and improving operations is why the collaboration between those two organizations is so important. Any policies or perceptions that undermine the importance of growing sales during a market constrained setting need to be dealt with and will be discussed in the next section.

A second simulation was performed in which the production system was the constraint and unable to provide a sufficient number of doses to meet the expected demand. Although the parameter ordering could have been anticipated by persons familiar with the Theory of Constraints, that method of analyzing the system had not previously been performed within Novartis and the relative importance needed to be determined.

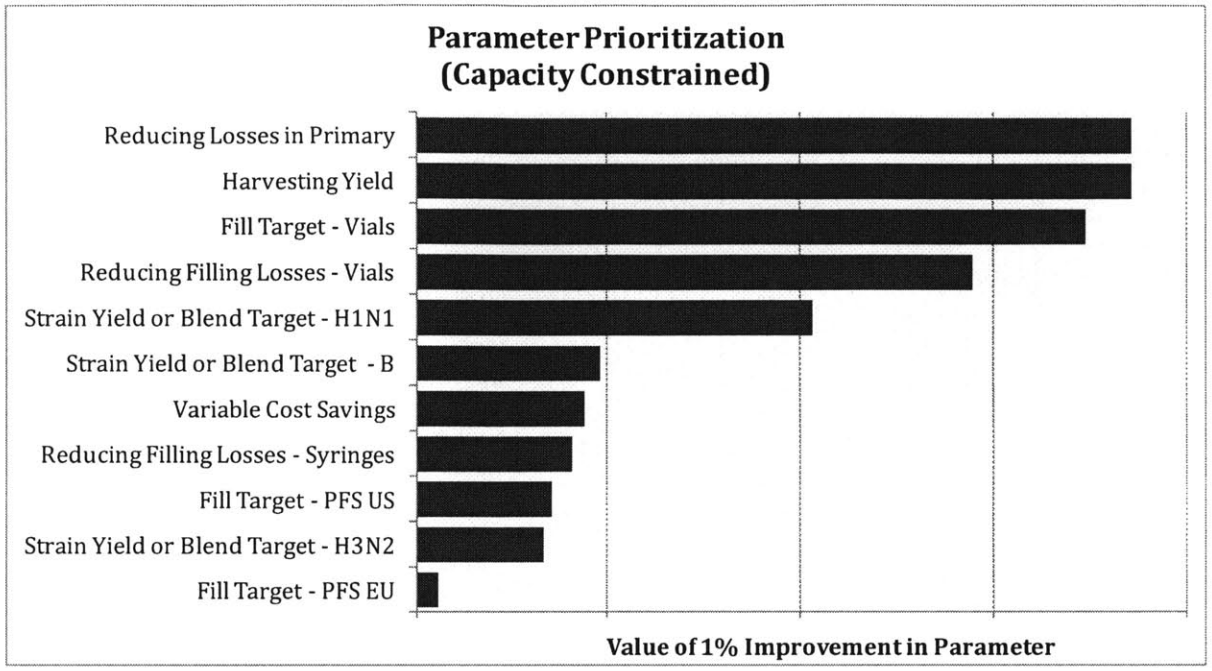

Figure 29: Parameter Prioritization During Capacity Constrained Season

As was explained in Chapter 4, the bottleneck of the process is at the transition between PIU and the incoming stage of primary upstream. Because the bottleneck is so early in the process, every parameter is either essentially increasing the number of doses through the bottleneck or reducing the amount of losses after it.

Reducing the primary losses is tremendously important because it saves material that has already passed through the bottleneck. Also, because it improves every strain it is the most effective at increasing the overall throughput rate and contribution margin. Additionally, harvesting yield also significantly increases throughput by increasing the utilization of the bottleneck. Any bad egg that is sent through the bottleneck is essentially wasting a spot in the constraint that can never be recovered. By increasing the harvesting yield, more good eggs are going through upstream, effectively increasing the bottleneck.

Improving the yields or blend targets for each of the strains has a different impact depending on the original values. By improving these parameters, more doses are able to fit through the bottleneck. Strains with low yields or high blend targets, are scheduled more frequently than the better performing strains. Therefore, improving the poor strains frees up more time in primary than improving the good strains would free up and that additional capacity can then be used to produce more batches.

Near the middle of the list is the value of reducing the variable costs by 1%. Reducing those costs during a capacity constrained season is significantly less productive than many of the other process improvements, yet was often treated as a higher priority in the past. This can be attributed to focusing too much on the cost side, rather than total profitability or because of the previous challenges of knowing the real profitability increases.

The final aspect that surprised managers and is an important insight because it differs from historical practice is where along the manufacturing process the important parameters lie. Traditional belief was that the further the product was in the manufacturing process, the more valuable it was. While this is technically true, the perceived magnitude of the value of the products later on in the process compared to those in the earlier stages was higher than the economic realities. Therefore, a disproportionate amount of effort was focused on the end of the process (reducing filling losses) at the expense of improving the earlier stages (reducing losses in primary). Reasons for this distorted perception and how to more accurately consider the value within the process will be discussed in a later section.

Managers from across the organization can now work together on a central model to understand which scenario they are operating in (a capacity constrained or market constrained season) and then prioritize their work appropriately.

In addition to knowing the value of the projects throughout the production process, the engineers who develop the seed virus often need to know how to plan their work. The seed team, which operates essentially as an R&D group, attempts to develop a good seed for each strain that will result in high yields. Seeds are created at the beginning of the year to start the new season's campaign, but there is often room for improvement later on. Managers need to decide if they should continue to use the original strains throughout the entire campaign or spend more to develop new seeds for increased yields later on in the year. Estimating the financial and customer benefit will help in determining if the additional costs are justifiable.

Figure 30 demonstrates the relationship between the yields of the strains and the contribution margin. It was generated by sweeping the amount of HA per egg for two of the strains and recording the contribution margin. At the beginning of the year, the baseline contribution margin would be at the point of the original strain yields. When a project to improve a seed is considered, the new condition would be transitioning to a new point on the curves with a new contribution margin. Because of the non-linear response of the contribution margin, this graph helps to show users in which scenarios a yield

improvement would be extremely valuable and when additional gains for a particular strain are minimal due to diminishing returns.

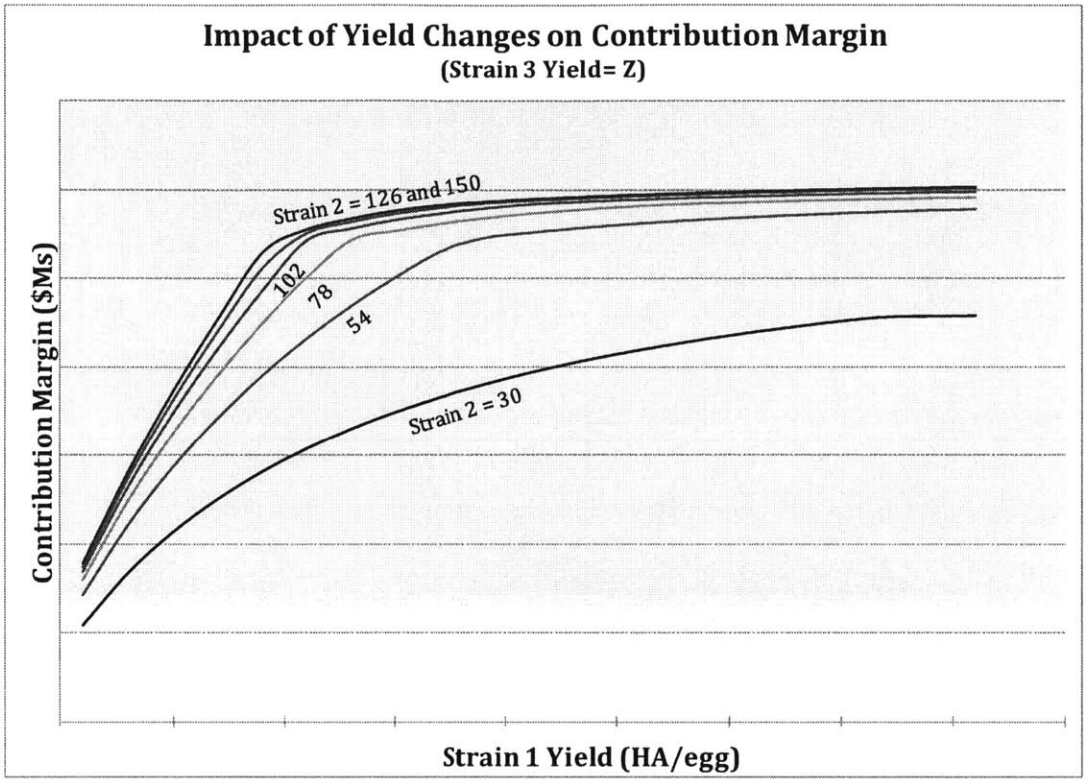

Figure 30: Impact of Strain Yields on Contribution Margins

7.1.4 Materials Purchasing

Flu vaccine production is heavily reliant on the decisions made when purchasing eggs, but it can be hard to strike the appropriate tradeoffs without knowing the rest of the system. Once modeled, we can see how to evaluate cost, quantity and quality.

First, because the exact percentage of good eggs is unknown and randomly varies each day, purchasing the right number of total eggs from the farms is a balance between the cost of buying too many (and having to discard the excess) and the opportunity cost of buying too few (and having less eggs than a full batch). Identifying the market condition to set the opportunity cost and combining it with the egg cost, average harvestable egg yield and standard deviation of egg yield in the newsboy model will determine how many extra eggs should be purchased in order to be optimal.

Eggs can also be negotiated based on perceived quality. A larger egg has more allantoic fluid, which can grow more HA, which will increase the amount of HA captured. In a capacity constrained season, this has the effect of increasing the bottleneck and can be tremendously valuable. Additionally, in a market constrained year, this still increases the batch size by only increasing part of the batch costs, so average costs are reduced. In either scenario, based on the data that maps egg size to the amount of HA, users can test how much the increase in egg quality is worth in a negotiation with the egg suppliers.

7.2 Short-term Recommendations

The inherent value of this project is not the immediate answer to a specific business problem; rather, it is improved methodology for addressing common production questions that appear in each campaign. As such, it is necessary to fully integrate the model and approaches into the regular discussions of the group. Incorporating the tool into the organization will help ensure that the shift from the current state to the desired state is achieved. Several steps were taken to help guide this transition and encourage full adoption, but additional effort is necessary to fully accomplish this.

7.2.1 Accounting Tools

Managers intuitively gravitated towards the new managerial accounting approach because of its ease of use. Instead of the complicated allocation scheme that was opaque to most, discussing the real impact of cost decisions was much more straightforward. Now, only two main questions are relevant for maintaining the cost information for each manager:

- If an additional dose is made at this step, what is the additional cost (marginal unit cost)?
- If an additional batch is made at this step, what is the additional cost (marginal batch cost)?

When evaluating a project, the additional questions are (Goldratt 1990, 45):

- What is the added investment of this project?
- What are the added operational expenses of this project?

For the tool to stay relevant, it is therefore essential that the information is always maintained in an accurate state. Additionally, steps must be taken to ensure that the model remains easy for managers to interact with. Therefore, it is important that the distinction between financial accounting and managerial accounting remains intact and not be blurred again.

7.2.2 Thinking Approach

Developing a systems thinking approach goes beyond simply the model, but the software can help managers understand the broader implications of their actions better than before. To truly encourage broader thinking, it is necessary to allow lots of exploration with the model from across the organization. Using the model will help identify unexpected results and allow them to dig into the details to understand why a particular behavior resulted and thereby help develop more trust in the tool.

Steps should be taken to make it easy for this type of discovery to take place, rather than limiting the number of users which can sometimes happen in enterprise software.

7.2.3 Integration of Objectives

Part of the value of having a model of the full organization is that it becomes very apparent when local objectives are coming at the expense of the global objective. Identifying where changes need to be made to bring the entire organization in line with the goal is an important step. Often, these suboptimal behaviors are a result of the unintended consequences of managerial policies.

For example, transfer prices and internal charges are necessary financial transactions for legal reasons, but can distort the behaviors of managers if they only look within their local site. These can lead to underinvestment in certain areas because of the local valuation based on the internal charges rather than market value. Additionally, the amount of risk can be distorted when one department has to buy the product from another at a high price, making the cost of doing business in an uncertain market appear higher than it really is.

Senior managers who want the entire system to perform at a globally optimal level must be able to identify where certain required transactions exist that lead to suboptimal behavior. Focusing on the appropriate system wide metrics by using the tool rather than historical local ones will encourage their managers to help overall rather than game local metrics.

7.2.4 Learning Experience

Even with the data provided by experimenting with the model, the organization still runs the risk of resorting to certain behaviors because of outdated mental models. Peter Senge warns of this, "...entrenched mental models will thwart changes that could come from systems thinking" (Senge 2006, 189). He also states, "The problem with mental models arise when they become implicit – when they exist below the level of our awareness" (Senge 2006, 166).

One such implicit mental model that was identified within the Novartis organization and which may potentially weaken the impact of the model, is the perception of a product's "value" in the manufacturing process. In fact, based on discussions with others, this belief seems common to many within the pharmaceutical industry, possibly leading to suboptimal behavior. It is not uncommon for entire industries to have a mental model that does not accurately reflect reality (Senge 2006, 166).

As the product moves through the manufacturing process, it is believed to become more valuable. Intuitively this seems obvious because more effort and materials have been invested throughout the process. (The phrase commonly used is that "more value has been added".) While this is technically true, what is distorted is the relative value at the different stages, thereby affecting decision making. Due to the long fully-coupled manufacturing process in biopharmaceuticals, often involving manual steps, product losses can occur. To be most effective, managers need to focus their quality control circles in the areas that will have the greatest gains by preventing these losses. The expected gains are based on the likelihood of success of the project, quantity of product saved and the value of the product that was saved. If managers have a distorted sense of the value along the process, they risk focusing in the wrong areas.

Existing mental models seem to be based on the intuition about value being added and on the way inventory is valued in the financial cost accounting system. In cost accounting, at each sub-step of the manufacturing process, the direct and allocated costs are added to the product to calculate its inventory value (for reporting reasons). Due to the large fixed costs being allocated to the product, the product value seems to rise steeply throughout the process. A manager looking at this form of valuation would focus almost exclusively at the very end of the process, where the product seems to be the most valuable. Focusing on minimizing losses at the end would appear to be the best way to improve the overall system.

Using this method can be misleading because allocations are merely an internal transfer of funds and do not reflect the real exchanges of money (Corbett 1998, 166). Instead of looking at the valuation approach for financial accounting, managers should evaluate how the system would respond to the question they are looking to answer: If fewer losses occurred at this particular stage in the process, what would be the economic benefit? Asking this question at each stage would show the relative value across the process and then guide the decision of where to focus efforts.

Results from this method are significantly different when thinking about the overall system. First, fixed costs become irrelevant because they are committed. Regardless of whether the losses occur at the very beginning of the process or the very end, those costs are going to be spent and therefore should not affect the perceived value of the product.

Secondly, the economic value of the product is not always simply the cost invested, but is a function of the production system and opportunity costs. When a product loss occurs before the bottleneck, the financial cost is the total variable cost to create the unit. Because the loss occurred prior to bottleneck, a replacement unit can still be produced and still generate the same revenue. If the system is market constrained (i.e. the market is the bottleneck) all losses are by definition before the bottleneck and fall into this first category. However, if the product loss occurs after the bottleneck, an additional unit cannot be made and revenue is lost. In that case, the economic cost of the loss was the price of the product minus any additional costs that would have needed to be spent to finish the product. The market situation and constrained resource determine where the transition occurs between losses based on costs spent and losses based on the opportunity cost.

Figure 31 shows a simple example in which the campaign year is capacity constrained to illustrate the concept. Inventory value based on financial accounting's approach of summing the variable costs with the allocated costs shows the increased value over time. It can be seen in the other curve how the alternative approach first tracks the variable costs and then climbs to an opportunity cost. A manager thinking in this new way realizes that the importance of preventing losses after the constraint is much higher than the old method would indicate. Additionally, the relative value of the product is significantly different in the two scenarios. When evaluating using cost accounting inventory valuations in this example, it appears that the final product is approximately three times as valuable as the product at the

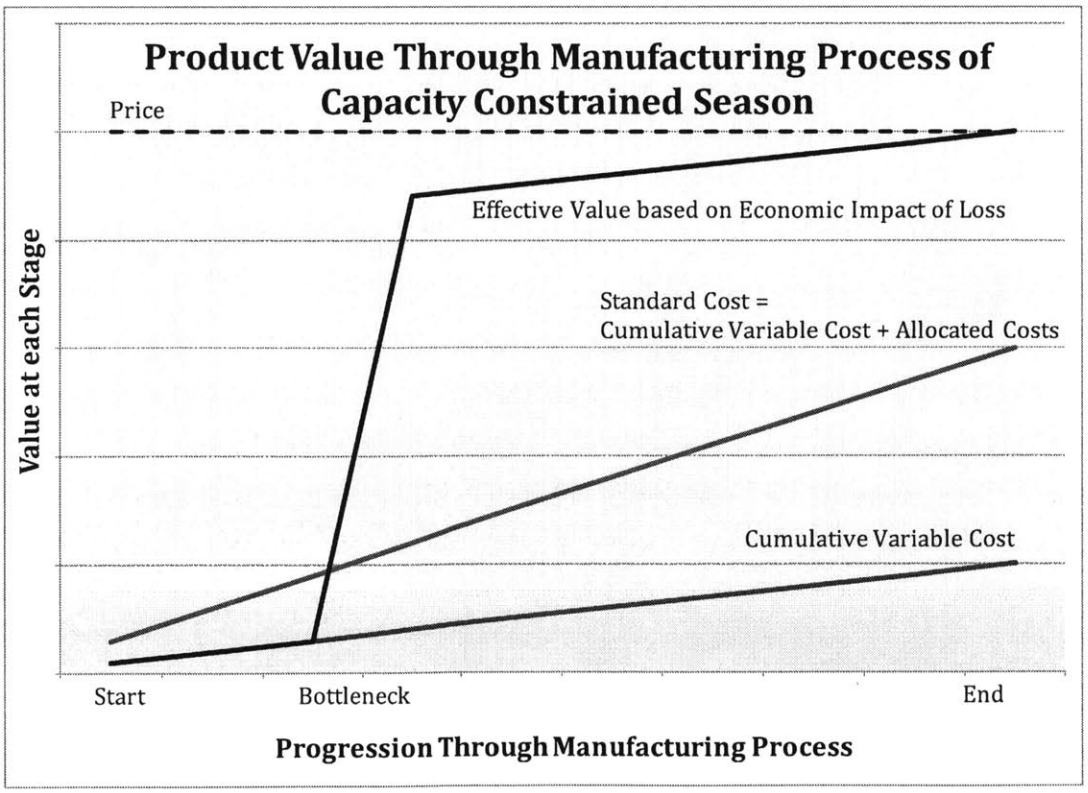

Figure 31: Valuation of Product During Capacity Constrained Season

bottleneck and thus a manager would likely focus on steps near the end of the process. However, in this new approach, it can be seen that the value after the bottleneck is much closer to the final product's value (about 90% of the full value). Even though the product is not quite as valuable at the bottleneck as it is at the end, they are quite close and it may be worthwhile to target projects in that area. A manager might very well arrive at a dramatically different conclusion on what areas are worth targeting. This provides a simple way of thinking about where losses are important, though the specific takeaways are dependent on the exact contour of the curve.

It was mentioned earlier in the conclusion that during a capacity constrained season, one of the most valuable improvements would be to reduce the losses in primary and that this was surprising to managers because of the perceived value at that stage compared to reducing losses in filling. Though confidential, the Fluvirin curve (when capacity constrained), is somewhat similar to Figure 31. We can see why thinking using the old valuation would focus at just the end for improvements, but now see that there are many other areas that can be improved upon to achieve a large gain.

As with the other recommendations, managers should experiment with the model and use it to question their existing mental models when the results seem surprising.

7.3 Long-term Recommendations

7.3.1 Importance of Flow

Throughout this thesis, several references have been made to the important differences between being market constrained and capacity constrained. Based on the data collection and interviews, it appears that the organization undervalued this distinction. While they often would work to increase throughput during times of being capacity constrained, the focus remained within operations even during market constrained seasons with the expectation of lowering costs through cost cutting. Instead, greater emphasis should be on production smoothing.

The distortion arises because of one of the previously discussed examples regarding standard costs. With the allocated fixed costs, standard costs made marginal production changes seem expensive. As a result, increasing production appears not to be very profitable, while reducing production appears to save a lot. With the real cost data, it was proven that increasing the production rate was more profitable than expected and reducing sales would be extremely costly.

A similar challenge exists at Toyota, a recognized operations leader, and much can be learned from them. Even though Toyota is known for its production line flexibility, experts acknowledge that its

cost structure creates an inherent risk. "... The system is extremely sensitive to fluctuations in the total volume of cars and trucks made. These types of shifts are very difficult to accommodate in a system which employees, because of job guarantees, are a fixed cost. So Toyota and other practitioners of lean production work hard at heijunka (production smoothing), in which the total volume the assembler manufactures is kept constant" (Womack, Jones, and Roos 2007, 154). To deal with this challenge, whenever demand has softened, Toyota has aggressively cut prices, utilized door-to-door car salesmen in Japan and even sent production workers out into the field as salesmen in order to keep production volumes steady (Womack, Jones, and Roos 2007, 66, 154, 190). With dedicated facilities, Novartis is prone to even more of this type of risk. Novartis should be just as aggressive during times of being market constrained and explore opportunities to develop new customers or find new markets.

One key difference between car and influenza vaccine manufacturing is the annual change in production volumes. Unlike a car factory which maintains approximately the same production volume each year, different strain yields each year can change the amount the site can produce. In years in which yields are higher than originally expected, an aggressive sales force may be able to help grow a new market to develop the necessary demand. The newly developed cost model can help guide their decisions to know what pricing strategies should be applied. When yields are low and they are capacity constrained, employees should work feverishly to increase throughput instead of trying to cut costs.

Continually focusing on the system constraint, whether in the market constrained season by developing new markets or in the capacity constrained year by improving production, will lead to gains for both Novartis and to its customers.

7.3.2 Implementing with Other Products

Finally, the techniques and insights from this project should be applied to other products and organizations at Novartis. A better understanding of the cost structure of all products will help managers identify the relevant costs to consider during decision making. When feasible, similar models should be developed to even further guide decision making and allow for experimentation. In the product families or sites that are simply too complicated to build an entire model in this way, insights can still be drawn from the Fluvirin project because many parallels exist in the cost structures and manufacturing processes.

As more techniques like this one facilitate learning about the various production systems throughout the business, the company will be better equipped to improve its performance and better serve its customers.

BIBLIOGRAPHY

Beckman, Sara, and Donald Rosenfield. 2008. *Operations Strategy : Competing in the 21st Century.* Boston: McGraw-Hill/Irwin.

Carnes, Kay, and Scott Hedin. 2005. Accounting for Lean Manufacturing: Another Missed Opportunity. *Management Accounting Quarterly*, Fall.

Corbett, Thomas. 1998. *Throughput Accounting : TOC's Management Accounting System.* Great Barrington MA: North River Press.

Cormen, Thomas. 2001. *Introduction to Algorithms.* 2nd ed. Cambridge Mass.: The MIT Press.

Danzon, Patricia, Nuno Pereira, and Sapna Tejwani. 2005. Vaccine Supply: A Cross-National Perspective. *Health Affais* 24, no. 3: 706-717.

Deming, W. Edwards. 1993. *The New Economics: for Industry, Government, Education.* 4th ed. Cambridge Mass.: Massachusetts Institute of Technology Center for Advanced Engineering Study.

———. 2000. *Out of the Crisis.* 1st ed. Cambridge Mass.: MIT Press.

Fine, Charles. 1998. *Clockspeed : Winning Industry Control in the Age of Temporary Advantage.* Reading Mass.: Perseus Books.

Garg, Ashish, Debashis Ghosh, James Hudick, and Chuen Nowacki. 2003. Roles and Practices in Management Accounting Today: Results from the 2003 IMA-E&Y Survey. *Strategic Finance*, July.

Goldratt, Eliyahu. 1983. Cost accounting is enemy number one of productivity. In *International Conference Proceedings, American Production and Inventory Control Society (APICS).* October.

———. 1990. *The Haystack Syndrome : Sifting Information out of the Data Ocean.* Croton-on-Hudson N.Y.: North River Press.

———. 1999. *Theory of Constraints : What is this thing called and how should it be implemented?* Croton-on-Hudson NY: North River Press.

HIDA. *2008-2009 Influenza Vaccine Production and Distribution: Market Brief.* Industry Publication. Health Industry Distributors Association. http://www.flusupplynews.com/documents/09_FluBrief_000.pdf.

Huntzinger, James. 2007. *Lean Cost Management : Accounting for Lean by Establishing Flow.* Ft. Lauderdale FL: J. Ross Pub.

Johnson, H, and Robert Kaplan. 1987. *Relevance Lost : The Rise and Fall of Management Accounting.* Boston Mass.: Harvard Business School Press.

Jones, Daniel. 2006. Heijunka: Leveling Production. *Manufacturing Engineering* 137, no. 2 (August): 29.

McNair, Carol, and Richard Vangermeersch. 1998. *Total Capacity Management : Optimizing at the Operational, Tactical, and Strategic Levels.* Boca Raton Fl.: St. Lucie Press.

Meadows, Donella. 1982. Whole Earth Models and Systems. *Co-Evolution Quarterly* (Summer): 98-108.

Noreen, Eric, Debra Smith, and James Mackey. 1995. *The Theory of Constraints and its Implications for Management Accounting*. Great Barrington, MA: North River Press.

Novartis. *Production cycle for annual influenza vaccine*. Internet Publication. http://www.novartisvaccines.com/products-diseases/influenza-products/influenza_vaccines_cycle/index.shtml.

———. 2008. About Novartis Vaccines - Our mission. *About Novartis Vaccines - Our mission*. August 8. http://www.novartisvaccines.com/about-vaccines/our_mission.shtml.

———. 2010a. *Influenza Virus Vaccine*. Vian and Syringe leaflet text. July 19. http://www.fda.gov/downloads/BiologicsBloodVaccines/Vaccines/ApprovedProducts/UCM123694.pdf.

———. 2010b. *Novartis Vaccines begins shipment of seasonal influenza vaccine Fluvirin to U.S. market to help protect ahead of 2010-2011 flu season*. Media Release. Cambridge, MA, July 29. http://www.novartisvaccines.com/newsroom/media-releases/2010/Fluvirin_Press_Release_7_29_10.pdf.

———. 2010c. Company History. *Company History*. December 15. http://www.novartis.com/about-novartis/company-history/index.shtml.

———. 2010d. The Novartis Mission. *The Novartis Mission*. December 21. http://www.novartis.com/about-novartis/our-mission/index.shtml.

Ohno, Taiichi. 1988. *Workplace Management*. Cambridge Mass.: Productivity Press.

Rother, Mike, and John Shook. 2003. *Learning to See : value stream mapping to create value and eliminate muda*. 1st ed. Brookline MA: Lean Enterprise Institute.

Sands, Neil. 2010. CSL Withdraws Child Flu Vaccine. *The Wall Street Journal*, June 1.

Senge, Peter. 2006. *The Fifth Discipline : The Art and Practice of the Learning Organization*. Rev. and updated. New York: Doubleday/Currency.

Stratton, William, Denis Desroches, Raef Lawson, and Toby Hatch. 2009. Activity-Based Costing: Is It Still Relevant. *Management Accounting Quarterly*, Spring.

Womack, James, and Daniel Jones. 2003. *Lean Thinking*. édition revue et mise à jour. New York: Free Press.

Womack, James, Daniel Jones, and Daniel Roos. 2007. *The Machine That Changed the World : The Story of Lean Production - Toyota's Secret Weapon in the Global Car Wars that is Revolutionizing World Industry*. [New ed.]. London: Simon & Schuster.

APPENDIX

Figure 32: Screenshot – Schedule (1)

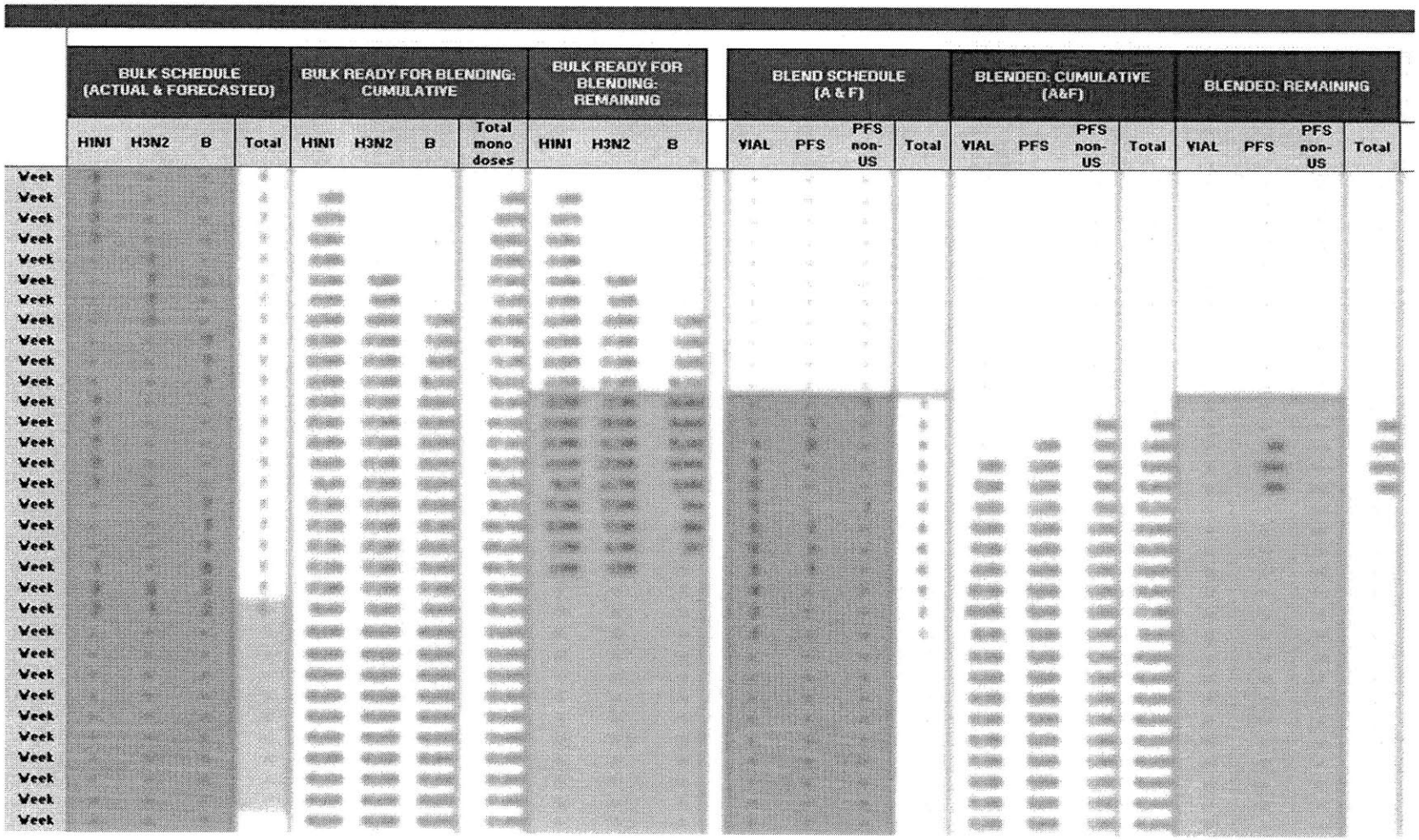

Figure 33: Screenshot – Schedule (2)

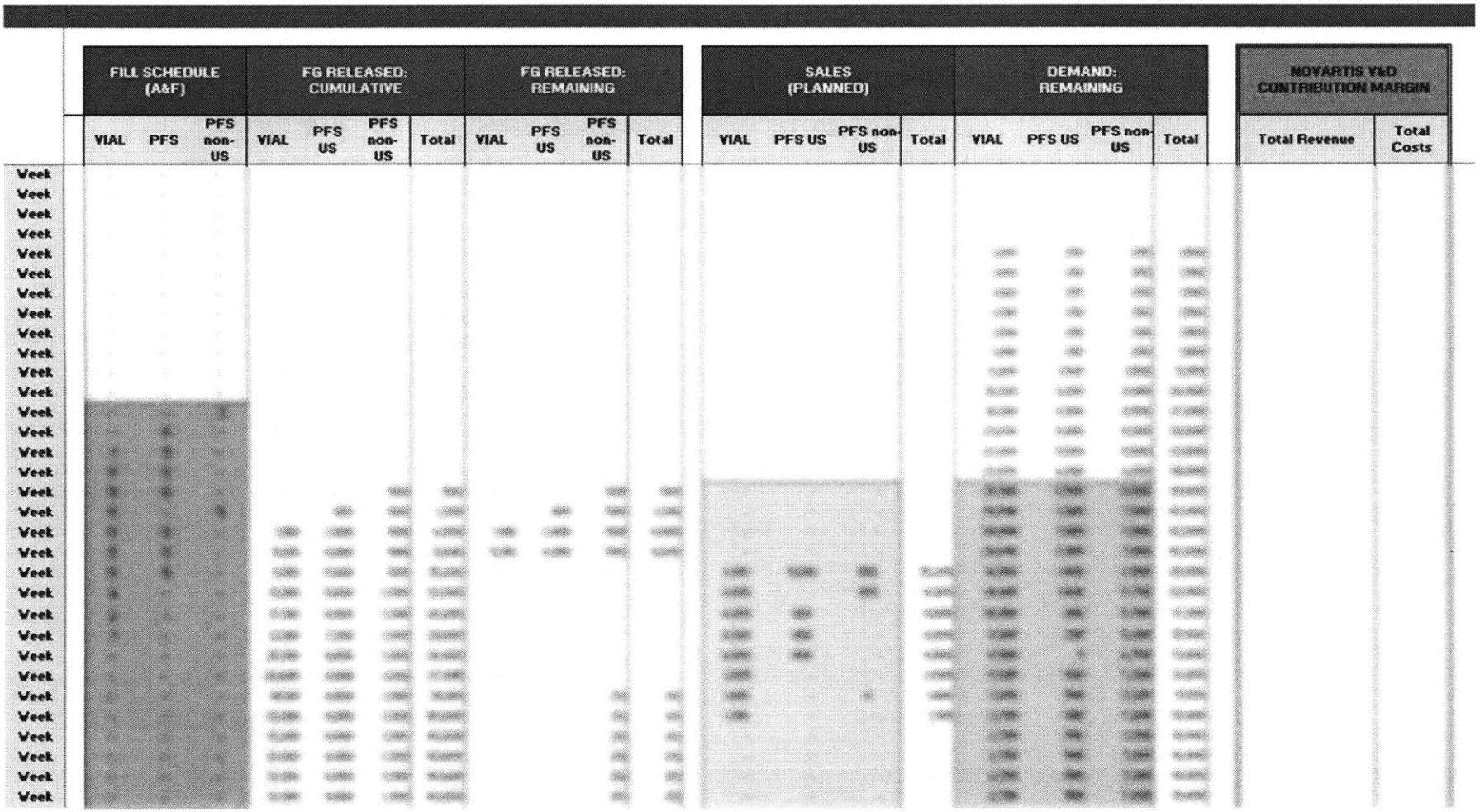

Figure 34: Screenshot – Primary

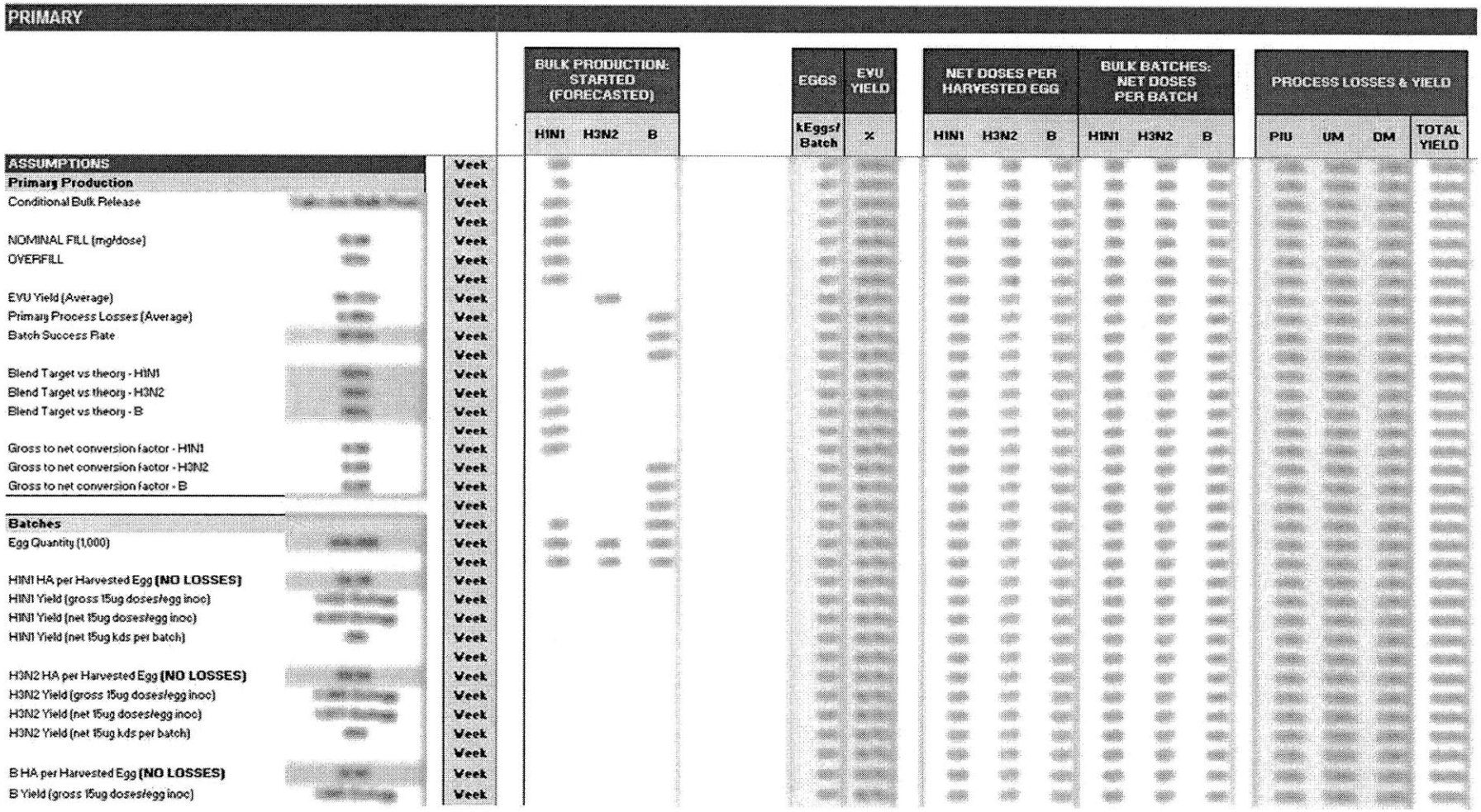

Figure 35: Screenshot – Blend

BLEND

ASSUMPTIONS	
Blending	
Batch Size - Vial	[illegible]
Batch Size - PFS US	[illegible]
Batch Size - PFS non-US	[illegible]
Blend ready for fill	[illegible]

	BLENDING PRODUCTION: STARTED				BLENDING PRODUCTION: COMPLETED			
	VIAL	PFS	PFS non-US	Cumulative	VIAL	PFS	PFS non-US	Total
Week								
Week								
Week								
Week								
Week								
Week								
Week								
Week								
Week								
Week								
Week			[illegible]	[illegible]				
Week		[illegible]		[illegible]			[illegible]	[illegible]
Week	[illegible]	[illegible]		[illegible]		[illegible]		[illegible]
Week	[illegible]			[illegible]	[illegible]	[illegible]		[illegible]
Week	[illegible]			[illegible]	[illegible]			[illegible]
Week	[illegible]		[illegible]	[illegible]	[illegible]			[illegible]
Week	[illegible]	[illegible]		[illegible]	[illegible]		[illegible]	[illegible]
Week	[illegible]	[illegible]		[illegible]	[illegible]	[illegible]		[illegible]
Week	[illegible]	[illegible]		[illegible]	[illegible]	[illegible]		[illegible]
Week	[illegible]			[illegible]	[illegible]	[illegible]		[illegible]
Week	[illegible]			[illegible]	[illegible]			[illegible]
Week	[illegible]			[illegible]	[illegible]			[illegible]
Week				[illegible]	[illegible]			[illegible]
Week				[illegible]				
Week				[illegible]				
Week				[illegible]				
Week				[illegible]				
Week				[illegible]				
Week				[illegible]				
Week				[illegible]				
Week				[illegible]				
Week				[illegible]				
Week				[illegible]				
Week				[illegible]				

Figure 36: Screenshot – Fill

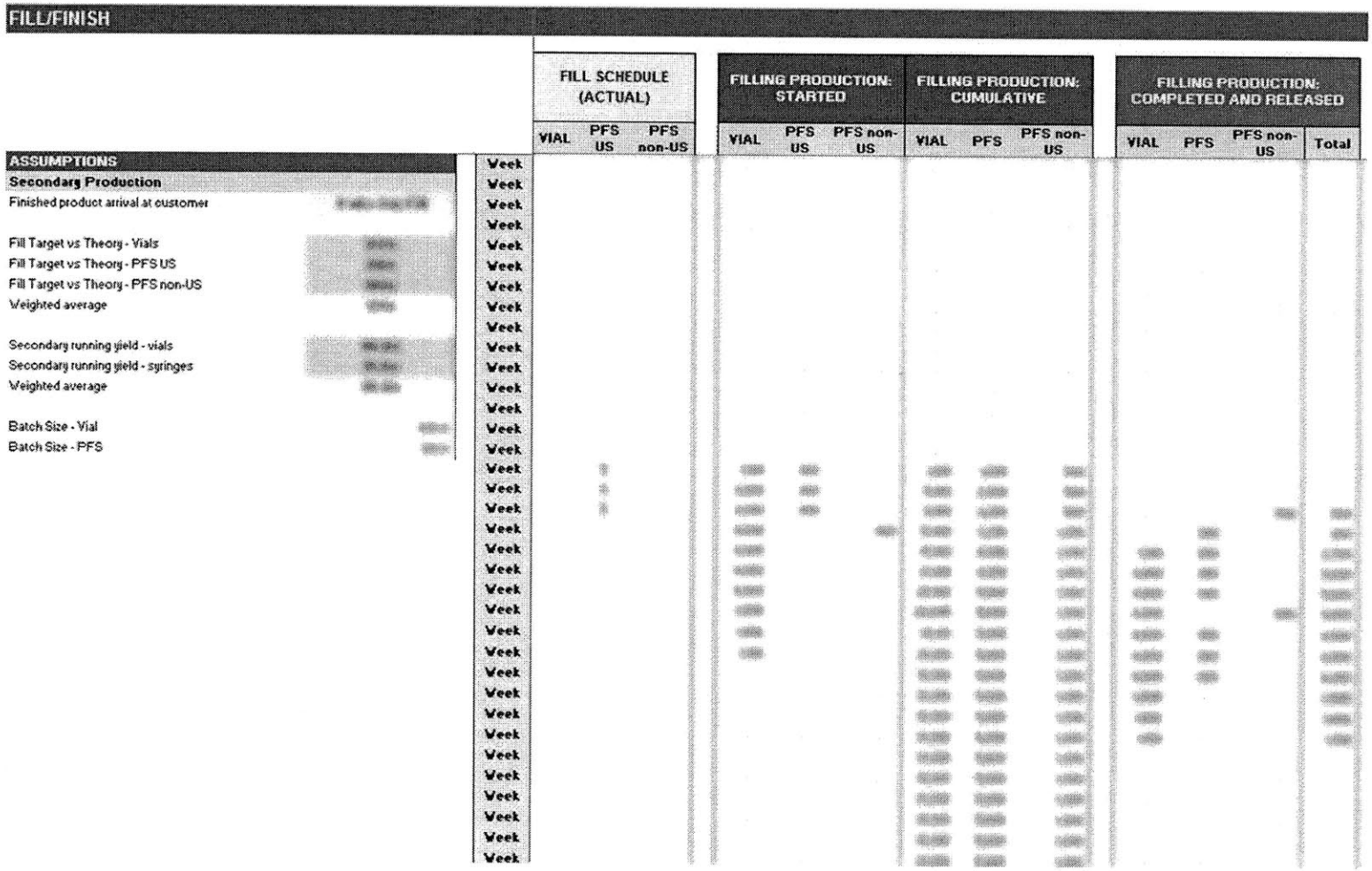

Figure 37: Screenshot – Market Demand/Delivery

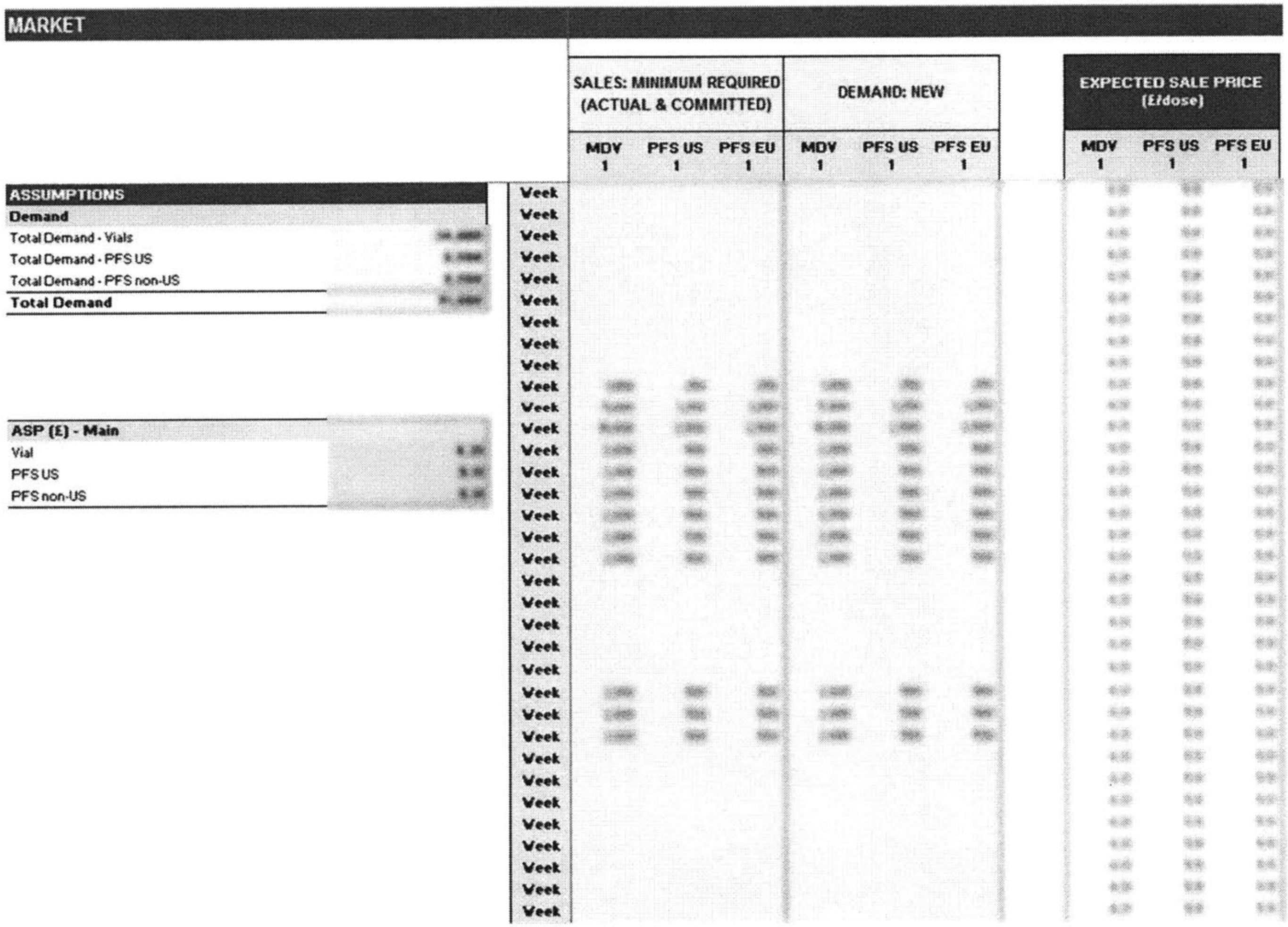

Figure 38: Screenshot – Cost Summary